Visual Perception of Solid Shape from Occluding Contours

Jack M. H. Beusmans
Dept. of Information and Computer Science
University of California, Irvine, CA 92717

Technical Report 90-40
November 15, 1990

This work was supported by ONR Contract N00014-88-K-0354 and NSF Grant IRI-8700924.

UNIVERSITY OF CALIFORNIA

IRVINE

Visual Perception of Solid Shape from Occluding Contours

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Information and Computer Science

by

Jacobus Maria Hubertus Beusmans

Dissertation Committee:

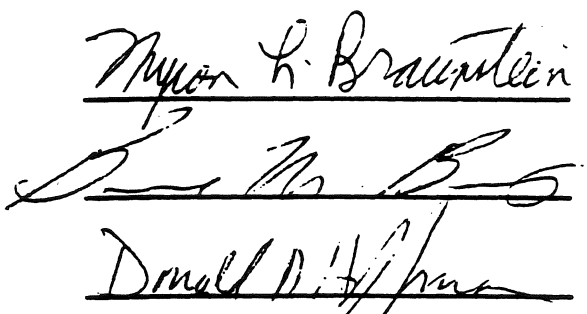
Professor Donald D. Hoffman, Chair
Professor Myron L. Braunstein
Professor Bruce M. Bennett

1990

The dissertation of Jacobus Maria Hubertus Beusmans is approved,

and is acceptable in quality and form for

publication on microfilm:



Committee Chair

University of California, Irvine

1990

aan mijn ouders,
Jan Beusmans & Tiny van Montfort

Table of contents

List of figures vi
List of tables viii
Acknowledgements ix
Curriculum vitae x
Abstract xii

Chapter 1. Overview and scope 1

Chapter 2. Visual perception and occluding contours:
A literature review 4
2.1 Occluding contours and the visual mapping 4
2.2 Occluding contours and the spherical mapping 10
2.3 Occluding contours and equidistance lines 13
2.4 Motion along occluding contours 15
2.5 Occluding contours and egomotion 20
2.6 Occlusion in depth and shape perception 21
2.7 Shape from occluding contours 27
2.8 Occluding contours and object recognition 28

Chapter 3. Depicting occluding contours 27
3.1 Contours rendered as silhouettes 27
3.2 Contours as discontinuities in light intensity 28
3.3 Contours as discontinuities in velocity fields 28
3.4 Contours as uncorrelated regions in stereo images 29

Chapter 4. Inferring object shape and motion from occluding contours . . 30
4.1 Computing object motion from occluding contours 34
4.2 Computing shape from occluding contours 48
4.3 Models of the perception of shape from contours 52
4.3.1 Rim prediction model 52
4.3.2 Tangent plane prediction model 54
4.3.3 Curvature model 54
4.4 Summary 55

Chapter 5. Visual perception of surface orientation 56

Chapter 6. Visual perception of solid shape from occluding contours . . . 70
6.1 Discriminating between solid and planar objects . . . 70
6.2 Perception of angular velocity of rotating ellipsoids . . . 91
6.3 Perception of shape from contours of rotating ellipsoids . . . 105
6.4 Summary . . . 112

Chapter 7. Summary and discussion . . . 114
7.1 Visual perception of solid shape and object motion from contours . . . 114
7.2 Visual perception of surface orientation from motion . . . 117
7.3 Visual perception of shape from contours and optic flow fields combined . . . 121

Bibliography . . . 123

Appendix A. Selected algorithms and proofs . . . 131
A.1 Coordinate systems . . . 131
A.2 Rotating ellipsoids . . . 132
A.3 Translating ellipsoids . . . 132
A.4 Rims and contours of ellipsoids . . . 133
A.5 Envelope of ellipse rotating in the plane . . . 135
A.6 Computing the differential invariants of the optic flow field . . . 135
A.7 Computing conjugation relationships from the slant field . . . 136
A.8 Velocity field induced by rotating objects . . . 138
A.9 Velocity field of a rotating plane . . . 139
A.10 Deformation of the optic flow of a cylinder translating in depth . . . 142
A.11 Deformation of optic flow of cylinder rotating about vertical length axis . . . 145
A.12 Deformation of optic flow of plane translating in frontoparallel direction . . . 146
A.13 Deformation of optic flow of plane rotating about frontoparallel axis . . . 147
A.14 Deformation of the optic flow of a plane translating in depth . . . 147

Appendix B. Discriminating ellipsoids and ellipses: raw data . . . 149

Appendix C. Glossary . . . 171

List of figures

2.1 Generic visual events . . . 5
2.2 Occluding contours of flying saucer . . . 6
2.3 Gaussian image of flying saucer . . . 11
2.4 Gaussian image of bean . . . 12
2.5 Gaussian images of rims and contours for flying saucer . . . 14
2.6 Equidistance lines on mannequin . . . 16
2.7 Stable forms of scalar field along singularities of 2D to 2D maps . . . 16
2.8 Equidistance lines and contours of bean . . . 17
2.9 Motion of a single sinusoidal grating and of plaid . . . 18
2.10 Motion discontinuity along contour with accretion and deletion . . . 22
2.11 Motion discontinuity along contour without accretion and deletion . . . 23
2.12 Blimps and ellipsoids . . . 24
2.13 Aspect of torus and visual potential of tetrahedron . . . 25
4.1 Dupin's indicatrices on sphere . . . 32
4.2 Slant field and conjugated directions . . . 33
4.3 Viewing cone . . . 35
4.4 Envelope of Gaussian images of rims . . . 37
4.5 Points of correspondence on contours of rotating ellipsoids . . . 39
4.6 Rims on sphere under perspective projection . . . 41
4.7 Envelopes of contours of rotating ellipsoids . . . 44
4.8 Bitangents of rotating objects . . . 47
4.9 Computation of curvature from dynamic contours . . . 49
4.10 Models of shape representation . . . 53
5.1 Deformation induced by receding cylinder . . . 58
5.2 Slant judgments for rotating planes . . . 64
5.3 Significance of slant judgments for rotating planes . . . 65
6.1 One-parameter family of ellipsoids . . . 71
6.2 Sequences of contours of rotating ellipsoids . . . 74
6.3 Influence of angular velocity on discrimination measures . . . 76
6.4 Absolute increase of discrimination measures under noise . . . 78
6.5 Relative increase of discrimination measures under noise . . . 79
6.6 Rim prediction measures . . . 80
6.7 Average results for discrimination experiments . . . 83
6.8 Psychometric curves based on curvature in direction of motion . . . 85
6.9 Contours of ellipsoids rotating about slanted axes . . . 90
6.10 Influence of angular velocity on discrimination measures . . . 92
6.11 Computation of angular velocity from occluding contours . . . 96
6.12 Angular velocity judgments for control stimuli . . . 98
6.13 Angular velocity judgments for contours and silhouettes . . . 99

6.14 Probe placements with contours of ellipsoids and ellipses 108
7.1 Shear and perception of depth 120
A.1 Spherical and cartesian coordinate systems 131
A.2 Conjugated directions from differences in slant 137
A.3 Rotating planes under orthographic and perspective projection 141
B.1 Psychometric curves based on various discrimination measures 162
C.1 Parabolic point and its indicatrix 172
C.2 Elliptic point and its indicatrix 173
C.3 Hyperbolic point and its indicatrix 174
C.4 Lines of curvature on an ellipsoid 174

List of tables

5.1 Divergence, curl, and deformation of rotating plane 61
5.2 Deformation of rotating and translating planes 67
6.1 Description of ellipsoids used in experiments B-1 through B-10 83
6.2 Discrimination results in experiments B-1 through B-8 86
6.3 Discrimination results with slanted axis of rotation 91
6.4 Angular velocity judgments 101
6.5 Probe placements with contours of ellipsoids and ellipses 111
A.1 Velocity field of rotating plane 143
A.2 Velocity field under orthographic and perspective projection 144
B.1 Results of experiment B.1 (actual angular velocity) 151
B.2 Results of experiment B.2 (actual angular velocity) 152
B.3 Results of experiment B.3 (actual angular velocity) 153
B.4 Results of experiment B.4 (actual angular velocity) 154
B.5 Results of experiment B.5 (actual angular velocity) 155
B.6 Results of experiment B.6 (actual angular velocity) 156
B.7 Results of experiment B.7 (actual angular velocity) 157
B.8 Results of experiment B.8 (actual angular velocity) 158
B.9 Results of experiment B.9 (actual angular velocity) 159
B.10 Results of experiment B.10 (actual angular velocity) 159
B.11 Results of experiment B.1 (estimated angular velocity) 160
B.12 Results of experiment B.3 (estimated angular velocity) 161

Acknowledgements

I would like to thank the members of my dissertation committee, Don Hoffman, Mike Braunstein, and Bruce Bennett, for their comments and suggestions. This work was supported by ONR Contract N00014-88-K-0354 and NSF Grant IRI-8700924.

Curriculum vitae

Jack M. H. Beusmans

Department of Information and Computer Science
University of California, Irvine
Irvine, CA 92717

Date of birth: February 9, 1958

Education

B.S.	University of Utrecht, The Netherlands (1979) Biomathematics
M.S.	University of Utrecht, The Netherlands (1983) Biomathematics
M.S.	University of California, Davis (1983) Plant Physiology
M.S.	University of California, Irvine (1985) Computer Science
Ph.D.	University of California, Irvine (1990) Computer Science
	Dissertation title: Visual Perception of Solid Shape from Occluding Contours
	Dissertation advisor: D. D. Hoffman

Academic positions

1990- Postdoctoral fellow, New York University

Professional affiliations

Association for Research in Vision and Ophthalmology
Institute of Electrical and Electronic Engineers

Publications

Beusmans, J. M. H. (in press). Computing occluding contours using spherical images. *Computer Vision, Graphics, and Image Processing.*

Beusmans, J. M. H., & Wieckert, K. E. (1989). Computing, research, and war: if "knowledge is power," where is responsibility? *Communications of the Association for Computing Machinery, 32*, 939–951.

Silk, W. K., & Beusmans, J. M. H. (1988) Mechanical properties within the growth zone of corn roots investigated by bending experiments. I. Preliminary observations. *American Journal of Botany, 75*, 990–995.

Beusmans, J. M. H., & Silk, W. K. (1988) Mechanical properties within the growth zone of corn roots investigated by bending experiments. II. Distribution of modulus and compliance in bending. *American Journal of Botany, 75*, 996–1002.

Beusmans, J. M. H. (1987). *Visual recognition of objects: behavioral, computational and neurobiological aspects* (Tech. Report No. 87-21). Irvine, University of California: Department of Information and Computer Science.

Beusmans, J. M. H., & Wieckert, K. E. (1987). Computing, research, and war: if "knowledge is power," where is responsibility? In *Proceedings Symposium on Directions and Implications of Advanced Computing* (pp. 158–166). Seattle, WA.

Beusmans, J. M. H., Hoffman, D. D., & Bennett, B. M. (1987). Description of solid shape and its inference from occluding contours. *Journal Optical Society of America A, 4,* 1155–1167.

Beusmans, J. M. H. (1983). *Production systems as a tool for integrating knowledge to predict protein structure. An artificial Intelligence approach.* (Tech. Report). University of Utrecht, The Netherlands: Bioinformatics Group.

Beusmans, J. M. H. (1983). *Young's modulus and stress relaxation correlated with growth pattern in corn roots.* Unpublished Master's thesis. Davis: University of California.

Beusmans, J. M. H. (1981). *Genetic and environmental regulation of the expression of the phoE gene in E. coli* (Tech. Report (in Dutch)). University of Utrecht, The Netherlands: Microbiology Group.

Abstract of the dissertation

Visual Perception of Solid Shape from Occluding Contours

by

Jacobus Maria Hubertus Beusmans

Doctor of Philosophy in Computer Science

University of California, Irvine, 1990

Professor Donald D. Hoffman, Chair

The relative motion of object and observer induces a motion field in the observer's visual image that is smooth everywhere except along the object's occluding contours. Thus, occluding contours and smooth motion fields can be viewed as complementary and as separate sources of information about an object's shape. I studied how the human visual system perceives solid shape from the occluding contours of rotating objects and from the smooth motion field induced by moving planar surface patches.

I propose a three-stage model for the perception of solid shape from the occluding contours of a rotating object. First, the object's motion is determined. I argue that this is only possible using points of correspondence and only when the object's axis of rotation is frontoparallel. In the second stage, the motion field along the contour is used to compute relative depth and surface curvature along the rim, the contour's pre-image. Third, local shape descriptors are propagated inside the figure to yield a global percept of solid shape. To determine which shape descriptors are computed by human subjects, I used a novel task in which subjects have to discriminate between flat ellipses and solid ellipsoids with varying thickness. I found that discriminability is proportional to the inverse of radial curvature but is not proportional to Gaussian or mean curvature. Certain slants of the axis of rotation decrease discriminability. Subjects who could discriminate ellipsoids and ellipses perceived the ellipsoids' angular velocity more veridically than did subjects who could not discriminate the two.

Any smooth motion field can locally be described by divergence, curl, and deformation. If the motion field is induced by a rotating plane, the amount of deformation is proportional to the plane's slant and its angular velocity. Similarly, for translating planes, deformation is proportional to slant and image motion. Slant judgments of human observers were to a first-order approximation proportional to deformation per se, that is, observers do not take object motion into account. Recent psychophysical evidence suggests that human subjects need motion discontinuities for this. Thus, contours might be necessary to correctly perceive slant from smooth motion fields.

Chapter 1
Overview and scope

I often say that when you can measure what you are speaking about,
and express it in numbers, you know something about it;
but when you cannot measure it, when you cannot express it in numbers,
your knowledge is of a meagre and unsatisfactory kind.

– Lord Kelvin (1883)

The subject of this dissertation is the human visual perception of solid shape. When looking at a solid object, human observers form a mental model which describes its shape; this model allows us to manipulate objects intelligently or to recognize them. The model is thought to be based on a combination of the many sources for shape information that are generally available in a natural image. Of the potential sources of shape information—such as stereo, shading, or motion parallax—I looked at occluding contours and at motion parallax. These two sources can be viewed as complementary in the sense that the visual image of an observer moving relative to a solid object has a motion field that is smooth everywhere except along the object's occluding contours. I studied how the human visual system perceives solid shape from the occluding contours of rotating objects and from the smooth motion field induced by moving planar surface patches.

An occluding contour can be characterized in many ways. It is the projection of a curve on an object that separates visible and invisible regions; as such it separates figure and ground in the image. Because an occluder must be closer to the observer than any objects it occludes, distance from the observer is discontinuous at occluding contours. As a consequence of this discontinuity, the optic flow field induced by a richly textured environment is also discontinuous at occluding contours. And as objects become visible or cease to be visible, texture elements appear and disappear at contours. Thus, occluding contours are prominent features of visual images and contain considerable information about the three-dimensional (3D) structure of the environment and the 3D shape of an object.

Looking at the occluding contours of a rotating solid, human observers can perceive solid shape (Todd, 1985). This observation, among others, challenged the longstanding belief that, in order to recover shape, successive views of an object must be projections of the same points on the object's surface (Marr, 1982; Wallach & O'Connell, 1953). This is not true for occluding contours. As an object rotates, its visible and invisible regions change continuously. Hence, successive occluding contours are not projections of the same points on the object.

I followed up Todd's work and studied in detail the human visual perception of shape from the occluding contours of rotating solid objects. I formulated a number of models for representing and computing solid shape from the contours of an object and tested them in

psychophysical experiments. These experiments confirm that human observers can perceive curved surfaces on the basis of occluding contours, that is, in the absence of correspondence. And the psychophysical data allow me to rule out certain models, but I was unable to narrow it down to only one of the models. As perceived shape and motion of an object are related, I measured both for rotating ellipsoids. Subjects who could not perceive solid shape, could not perceive the correct angular velocity; their response was based on 2D image velocities. Subjects able to perceive solid shape could perceive angular velocity correctly for certain renditions of the occluding contours but not for others.

The remainder of the dissertation is organized as follows. In **Chapter 2**, I review the theoretical results regarding occluding contours, namely what shape information can theoretically be derived from them. I also review some psychophysical experiments on the role of contours in shape perception and the recognition of objects.

If we want to study the role of contours in visual perception, we need to design visual stimuli that exclude, as much as possible, other sources of shape information. In **Chapter 3**, I discuss the problem of depicting occluding contours of a solid object.

In **Chapter 4**, I outline several models for inferring solid shape from contours. The two principal models are the relative depth and the curvature model; the others are variations of these two. In the relative depth model, relative depth along the pre-image of the contour is computed from the instantaneous velocity field along the contour; relative depth measurements from successive views are combined to yield a percept of solid shape. In the curvature model, surface curvature is computed directly from the deforming contours; surface curvature along the contour's pre-image is combined to obtain a percept of solid shape.

Chapter 5 is an intermezzo on the perception of surface orientation from motion parallax. Any smooth motion field can locally be described by three differential invariants, namely divergence, curl, and deformation. If the motion field is induced by a rotating plane, the amount of deformation is proportional to the plane's slant and its angular velocity. Similarly, for translating planes, deformation is proportional to slant and image motion. I found that slant judgments of human observers are to a first-order approximation proportional to deformation per se, that is, observers do not take object motion into account. Recent psychophysical evidence suggests that human subjects need motion discontinuities for this. In other words, occluding contours might be necessary to correctly perceive slant from smooth motion fields.

In **Chapter 6**, I report on three psychophysical experiments. The first experiment was designed to determine which, if any, of the models outlined in Chapter 4 pertain to the human perception of occluding contours. The idea is to measure how well observers can discriminate between solid, volume-containing ellipsoids and planar ellipses as a function of the ellipsoids' thicknesses. In the second experiment, I measured the perceived angular velocity and the perceived shape of rotating ellipsoids as a function of their thicknesses. And in the third experiment, I used a dynamic probe dot to measure perceived depth inside the contours of a rotating ellipsoid, again as a function of the ellipsoid's thickness.

Chapter 7 summarizes and discusses the contributions of the work presented in this dissertation.

Appendix A contains various algorithms and proofs that were too lengthy or distracting to be included in the main text.

Appendix B contains detailed descriptions and the raw data of experiment 6.1.

Appendix C contains definitions and intuitive descriptions of mainly mathematical concepts, in particular concepts from differential geometry.

Chapter 2
Visual perception and occluding contours: A literature review

In this section, I summarize the results of theoretical and psychophysical investigations of occluding contours. I start with the role of occluding contours in the visual mapping and the spherical mapping; then I discuss the relationship between the depth map and contours, and the use of occluding contours in computing egomotion. I conclude the discussion of theoretical results with the inference of shape from occluding contours. As far as the psychophysical aspects of contours are concerned, I review the studies on the relationship between occlusion and depth perception, the perception of solid shape from the silhouette of a rotating solid, and on the role of occluding contours in object recognition.

2.1 Occluding contours and the visual mapping

Consider an opaque solid object and an observer at some distance. The *visual mapping* or projection, $\pi : \Re^2 \rightarrow \Re^2$, maps points on the surface of the object onto a two-dimensional (2D) image surface associated with the vantage point of the observer. Because solid objects enclose a finite volume of 3D space, π cannot be everywhere one-to-one. π is locally two-to-one along *occluding rims*, that is, curves on the object's surface that separate visible from invisible regions. This is the so-called fold singularity of π. Occluding rims project onto *occluding contours* in the image. At isolated points along an occluding rim, namely where the rim itself becomes invisible due to local occlusion, π can be three-to-one. This gives rise to a *cusp* along the occluding contour. Folds and cusps are the only stable singularities for mappings such as π (Whitney, 1955). Occluding contours circumscribe the image of an object, that is, they separate figure and ground.

If the observer moves around the object, the occluding rims and hence the occluding contours change smoothly except for an occasional abrupt qualitative change. These abrupt changes are called *visual events*; all the generic visual events are listed in Figure 2.1 (Arnold, 1984; Bruce, 1984a, 1984b; Koenderink, 1984b, 1987; Koenderink & van Doorn, 1976a; Rieger, 1987).

The direction of the rim and the viewing direction are related in an interesting manner. Let $d\mathbf{x}$ be the viewing direction and $\delta\mathbf{x}$ the direction of the rim at some point on the rim and in a local cartesian coordinate system; let $\delta\mathbf{N}$ be the change of surface normal in the direction of $\delta\mathbf{x}$. Then $(d\mathbf{x} \cdot \delta\mathbf{N}) = 0$, that is, the viewing direction and the direction of the rim are *conjugate* (Beusmans, Hoffman, & Bennett, 1987; Koenderink, 1984b; Koenderink & van Doorn, 1975, 1976b, 1978). Intuitively, the difference vector $\delta\mathbf{N}$ between surface normals at neighboring points on the rim has to be perpendicular to the local line of sight because the normals themselves are perpendicular to the line of sight.

For every non-asymptotic direction (in other words, for every direction in which normal curvature is not zero), there is a unique conjugate direction (Lipschutz, 1969). Because the

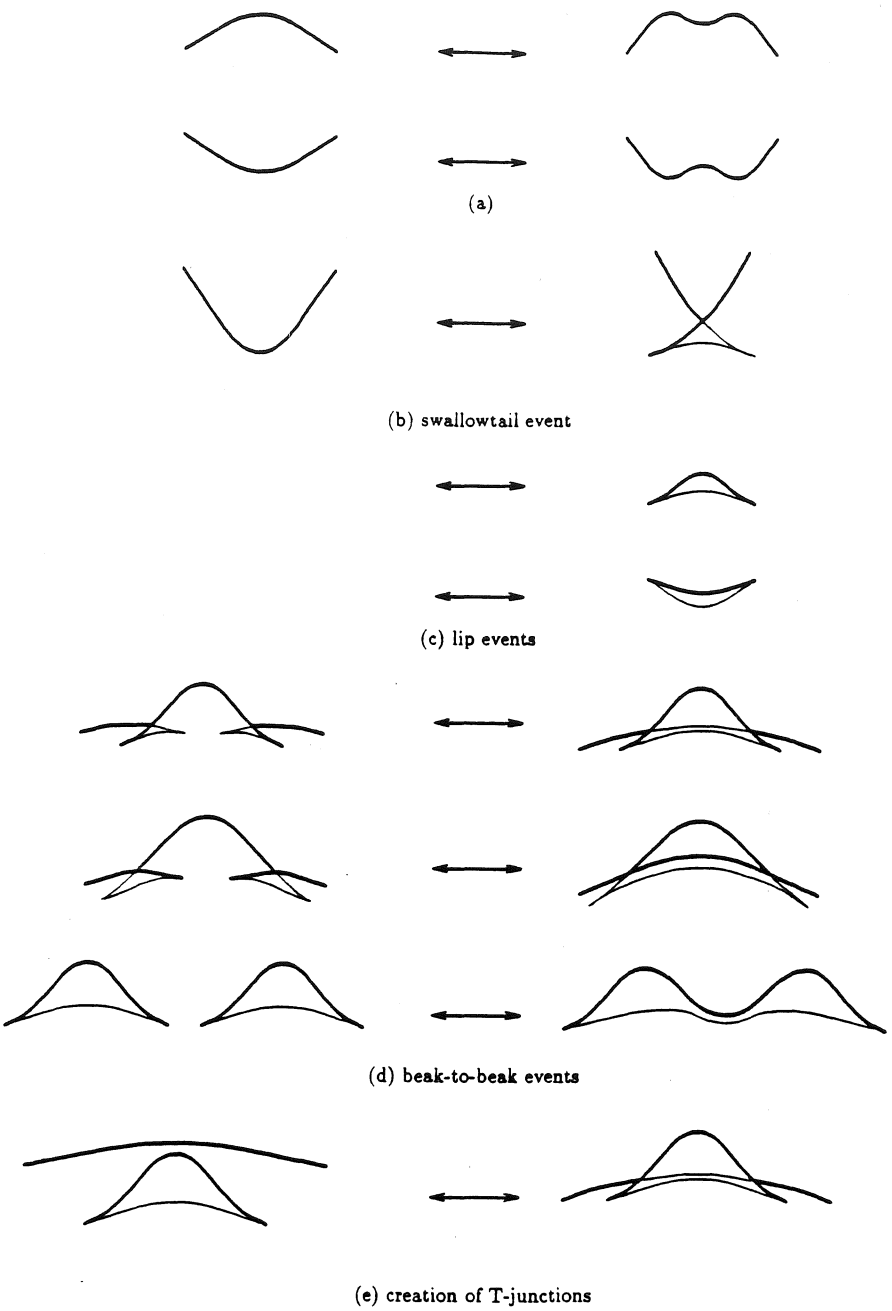


Figure 2.1 The generic changes in an occluding contour:
(a) a concave (convex) arc appears or disappears within a convex (concave) arc, creating or annihilating two inflection points;
(b) a concave arc occludes itself, or a T-junction and cusp disappear (swallowtail event);
(c) a hump or a dent appears or disappears (lips event);
(d) a fold splits off a new fold, or two folds coalesce (beak-to-beak event); two cusps and two inflections (dis)appear;
(e) two T-junctions appear or disappear (tangent crossing).
(from Beusmans, in press).

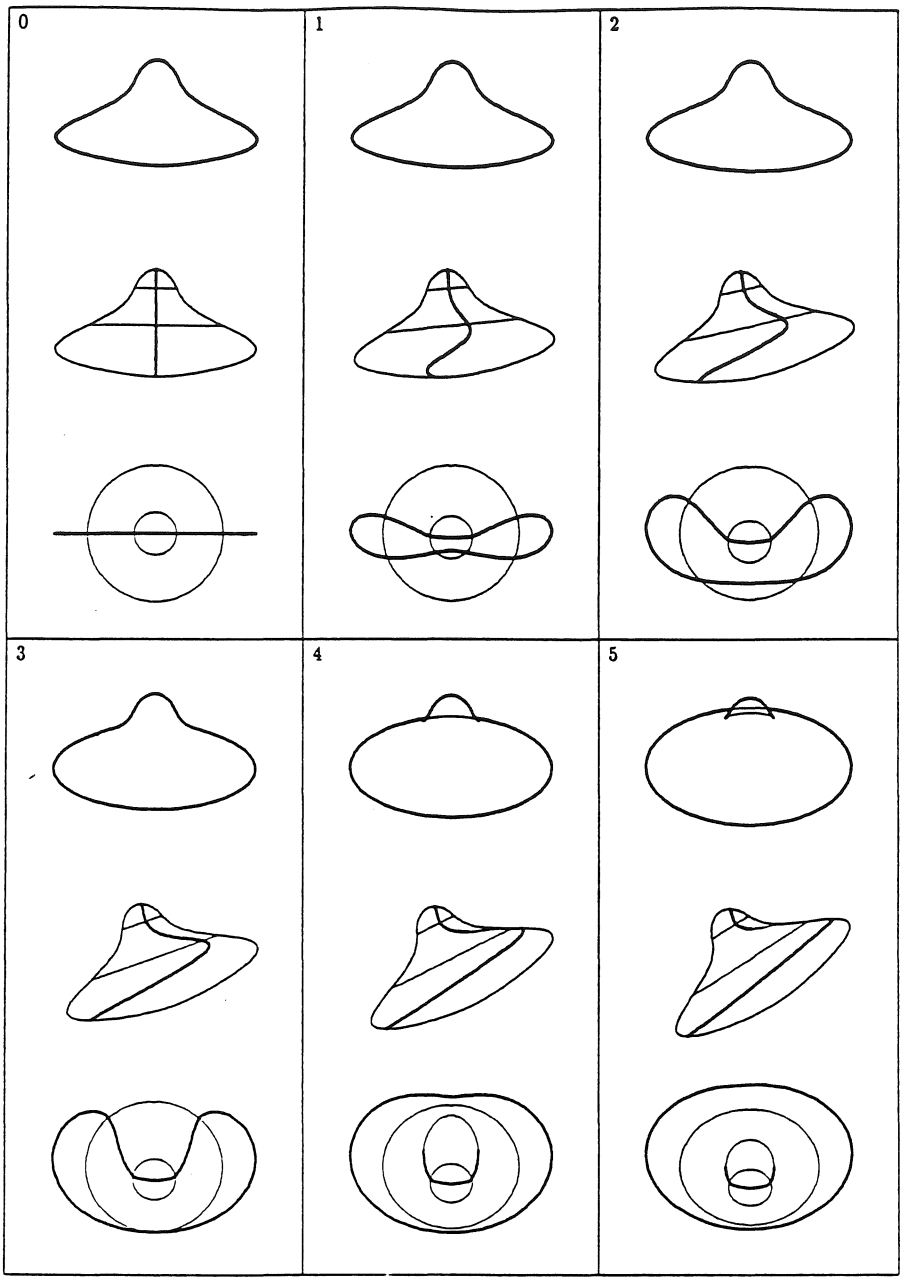


Figure 2.2 (a) Twelve views of the flying saucer shown in Figure 2.3 as it rotates at 6.6°/frame about the horizontal with its hump coming towards the viewer. In each frame is shown: (top) a frontal view, with invisible parts of the contours depicted with thin lines; (middle) a side view of parabolic curves and folds; and (bottom) a top view of folds and parabolic curves. Note that in the middle and bottom panels thick (thin) lines denote fold points with positive (negative) radial curvature. Between frames 3 and 4 a beak-to-beak event occurred; between 9 and 10 a lip event occurred. (b) Five intermediary views between 3 and 4 (3.0 is the same as 3 in (a); rotation is 1.1°/frame)
(from Beusmans, in press).

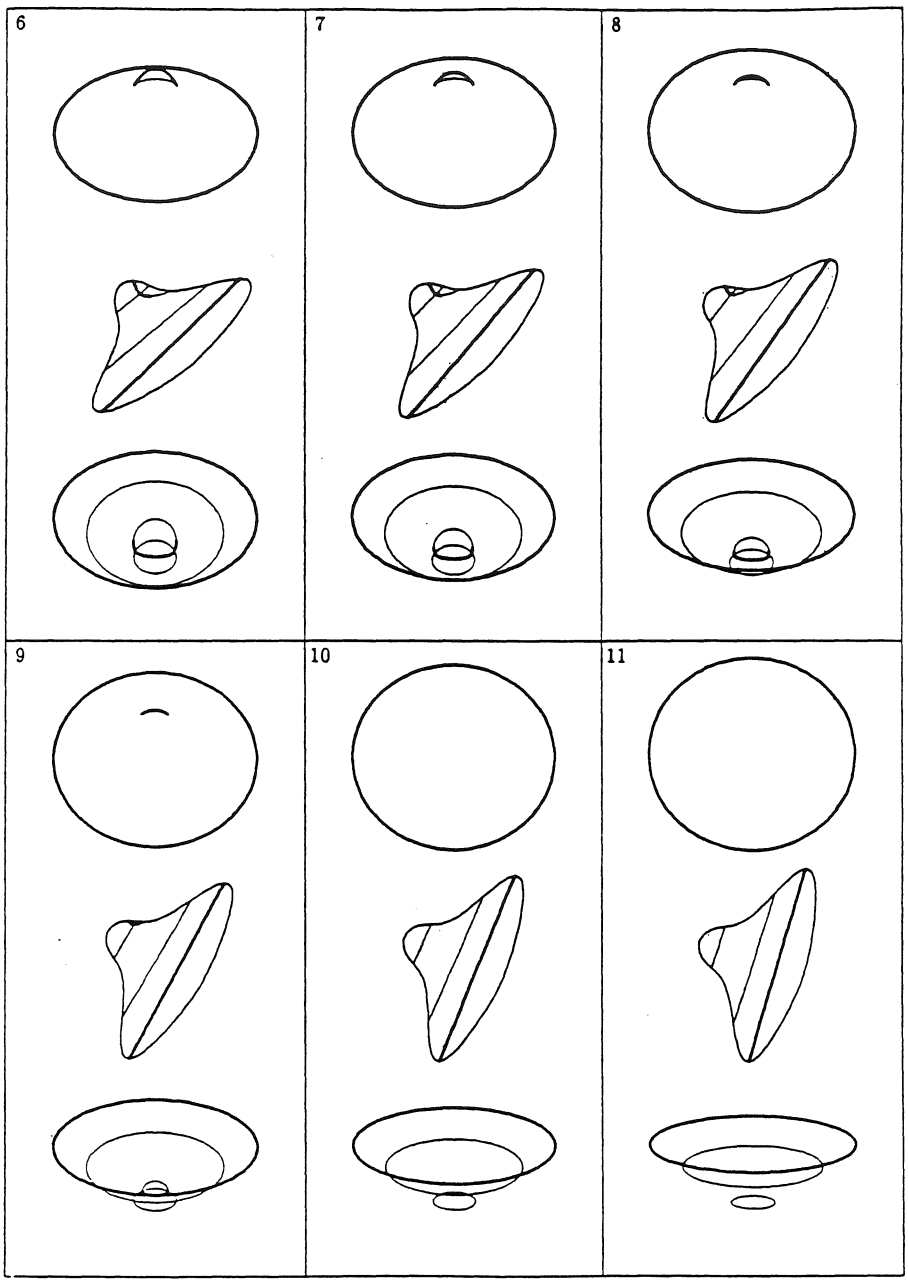


Figure 2.2a cont.

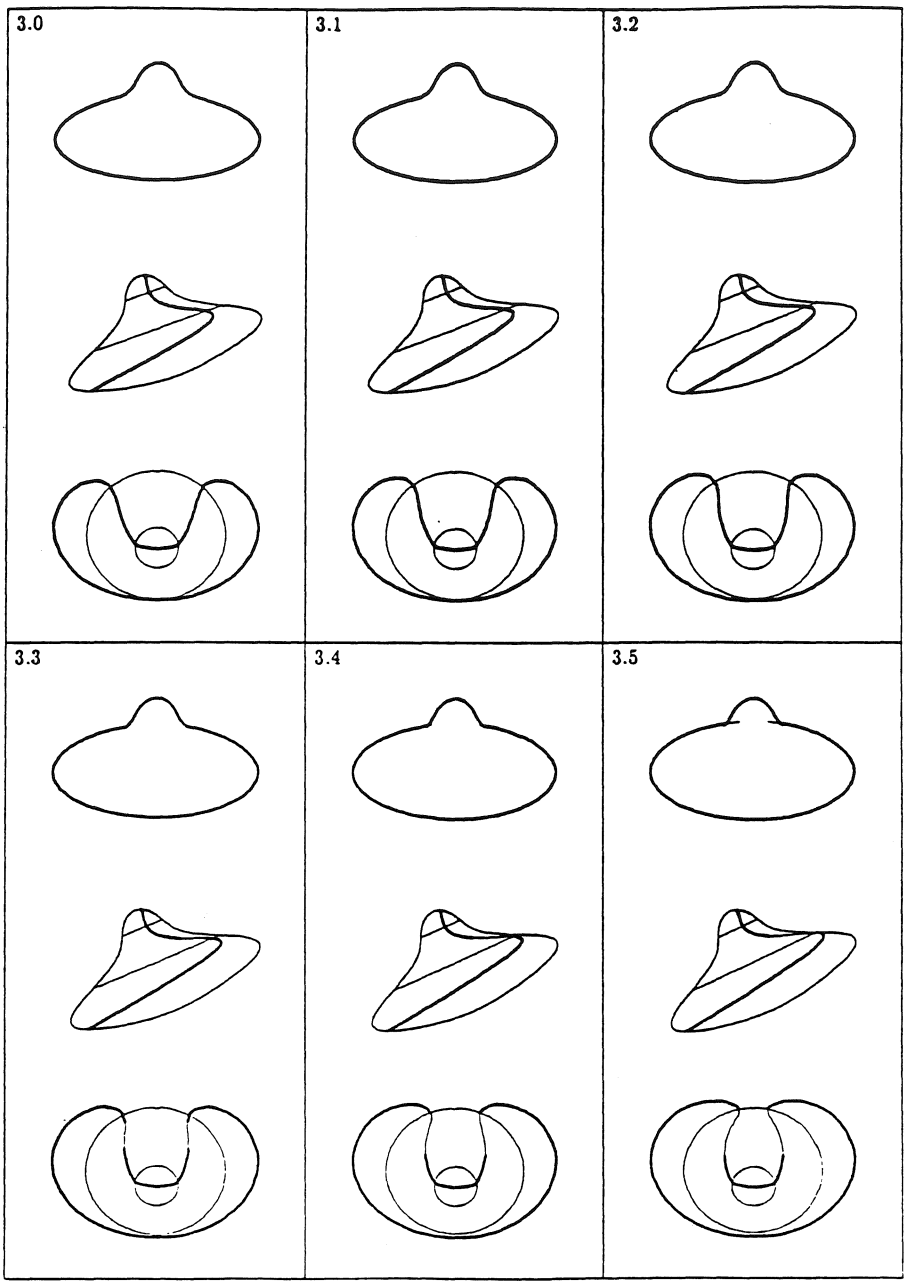


Figure 2.2b.

visual direction at any point on the rim cannot be asymptotic (there would not be a rim otherwise), it follows that for every visual direction there is only one direction in which the rim can go. This implies that rims do not intersect transversally.

The above remarks about visual events and the conjugation relationship between rim and visual direction are illustrated in Figure 2.2. Figure 2.2a shows a sequence of 12 views of the flying saucer depicted in Figure 2.3a as it rotates at 6.6°/view about the horizontal, its hump coming towards the viewer; Figure 2.2b shows 6 additional views between views 3 and 4 of Figure 2.2a (successive vantage points differ only 1.1°). The top row of each panel shows the occluding contours, with visible parts rendered by a thick curve, invisible parts by a thin one; the middle row shows a side view of the rims on the saucer's surface and its parabolic curves; and the bottom row gives a top view of the rims and parabolic curves only.

The following facts about rims and occluding contours and rims are demonstrated:

- Rims are rarely planar for generic objects; only views 0, 10, and 11 in Figure 2.2a have almost planar rims. The rims of special objects such as ellipsoids and spheres are always planar; interestingly, this is true under orthographic as well as perspective projection (section A.4).
- A rim crosses a parabolic line in the direction of zero normal curvature. (I mentioned above that the visual direction and the direction of the rim are conjugate. At parabolic points, every direction is conjugate to the direction of zero normal curvature because in that direction $\delta\mathbf{N} = 0$, so that $(d\mathbf{x} \cdot \delta\mathbf{N}) = 0$ for all $d\mathbf{x}$.) As the saucer happens to be a surface of revolution, the direction of zero normal curvature is everywhere perpendicular to the parabolic curve itself; thus, parabolic curves and rim cross at right angles (e.g., Figure 2.2a, view 3).
- Rims "cling" to highly curved regions of an object's surface but "sweep" over its flat regions. (Because the rim is defined as consisting of those points on the object's surface whose normal is perpendicular to the current viewing direction, the way a rim moves over the surface as the vantage point changes is determined completely by changes in the local surface normal, that is, local surface curvature. Thus, a highly curved region can accommodate larger changes in vantage point than a flat region. In other words, the rim tends to stay longer in curved regions than in flat ones.) Thus the larger of the two rims hardly changes position between views 6 and 11, differing 33° in viewing position, as it passes through a highly curved region of the saucer. In contrast, both rims change considerably between views 4 and 5 as they move in opposite directions away from the parabolic curve.
- Rims and cusps are stable singularities; changes in viewing position generally do not make them appear or disappear, but instead cause them to move over the object's surface.
- Occasionally rims do suffer abrupt changes:
  - Between views 3.2 and 3.3 in Figure 2.2b, two swallowtail events occurred, each generating two cusps and a T-junction which have become more obvious in view 3.5.
  - Between views 3.5 and 4, a beak-to-beak event occurred. Note that the viewing direction is parallel to the asymptotic direction of a parabolic point as this event occurs; the same happens in the next event.
  - Between views 9 and 10, a lips event occurred.

2.2 Occluding contours and the spherical mapping

The *spherical* or *Gauss mapping* relates points on the surface of an object with points on a unit sphere representing all possible surface normals. The spherical image of a point P is the unit surface normal at P as represented by a point on the unit sphere. As with the visual mapping, the only singularities of the spherical mapping are the fold and cusp. (To avoid possible confusion, it is worth stressing that the terms fold and cusp are generic and refer to certain singularity types in any mapping. In the present paper, we consider two mappings, viz. the visual mapping and the spherical mapping, each having its own distinct folds and cusps). Folds of the Gauss map occur at parabolic curves because one of the lines of curvature has an inflection here; if we were to trace this line of curvature, we would note that the direction of traversal on the unit sphere reverses upon crossing the parabolic line. This point of reversal is a singularity because the spherical mapping is locally 2-to-1. Cusps occur at isolated points along parabolic curves; at these so-called pedal points (Koenderink, 1984a, 1987), the direction of the parabolic line coincides with that of the line of curvature having the inflection.

These observations are illustrated in Figures 2.3 and 2.4. Figure 2.3a shows a stereoview of a "flying saucer," a surface of revolution consisting of a convex hump on top of a flattened ellipsoid, and separated from it by a region of saddle points (these and subsequent stereo views need to be free fused). The two thick curves labeled p_1 and p_2 are the two parabolic curves. Figure 2.3b depicts some relevant features of the saucer's spherical image. $S(p_1)$ and $S(p_2)$ are the spherical images of p_1 and p_2, resp.; both happen to be circles of latitude on the unit sphere. Figure 2.3c shows the path on the sphere that results from moving along the meridian connecting Q_4 and Q_1 on the saucer's surface; note that in crossing p_1 and p_2 the sign of normal curvature changes and with it the direction of traversal on the sphere at $S(p_1)$ and $S(p_2)$, resp. As a result the annulus on the unit sphere bounded by $S(p_1)$ and $S(p_2)$ is triply covered, each point having three pre-image points. For example, $S(Q_i)$ is the spherical image of points Q_1, Q_2, and Q_3, since all three have the same surface normal (Figure 2.3c). In contrast, all points outside the annulus, such as $S(Q_4)$, are covered only once. Note that the pre-image of a singly covered region always has positive Gaussian curvature, and that the three sheets of a triply covered region derive from regions with positive, negative, and again positive Gaussian curvature (going from the inside towards the outside).

Figure 2.4a shows a stereoview of a bean-like object, generated by bending an ellipsoid along its major axis, with its spherical image in Figure 2.4b. Now there is a single parabolic curve p; again, $S(Q_i)$ is triply covered. In Figure 2.4c, the lines of curvature of the patch of saddle points enclosed by p are represented schematically. Because we are considering a patch with negative Gaussian curvature the lines of curvature form a locally orthogonal net (this is true for any patch which does not include umbilical points where every direction is a principal direction; a sphere consists entirely of umbilical points). Hence, the closed parabolic curve p has at least two pedal points (P_0 and P_1), where its direction coincides with the local direction of zero (principal) normal curvature. The spherical image of the parabolic curve, shown schematically in Figure 2.4d, has cusps at $S(P_0)$ and $S(P_1)$. To understand intuitively why the spherical image has cusps at these locations, note that the direction of the parabolic curve at the pedal points is perpendicular to the direction of its spherical image. This is the case because an asymptotic direction and its spherical image

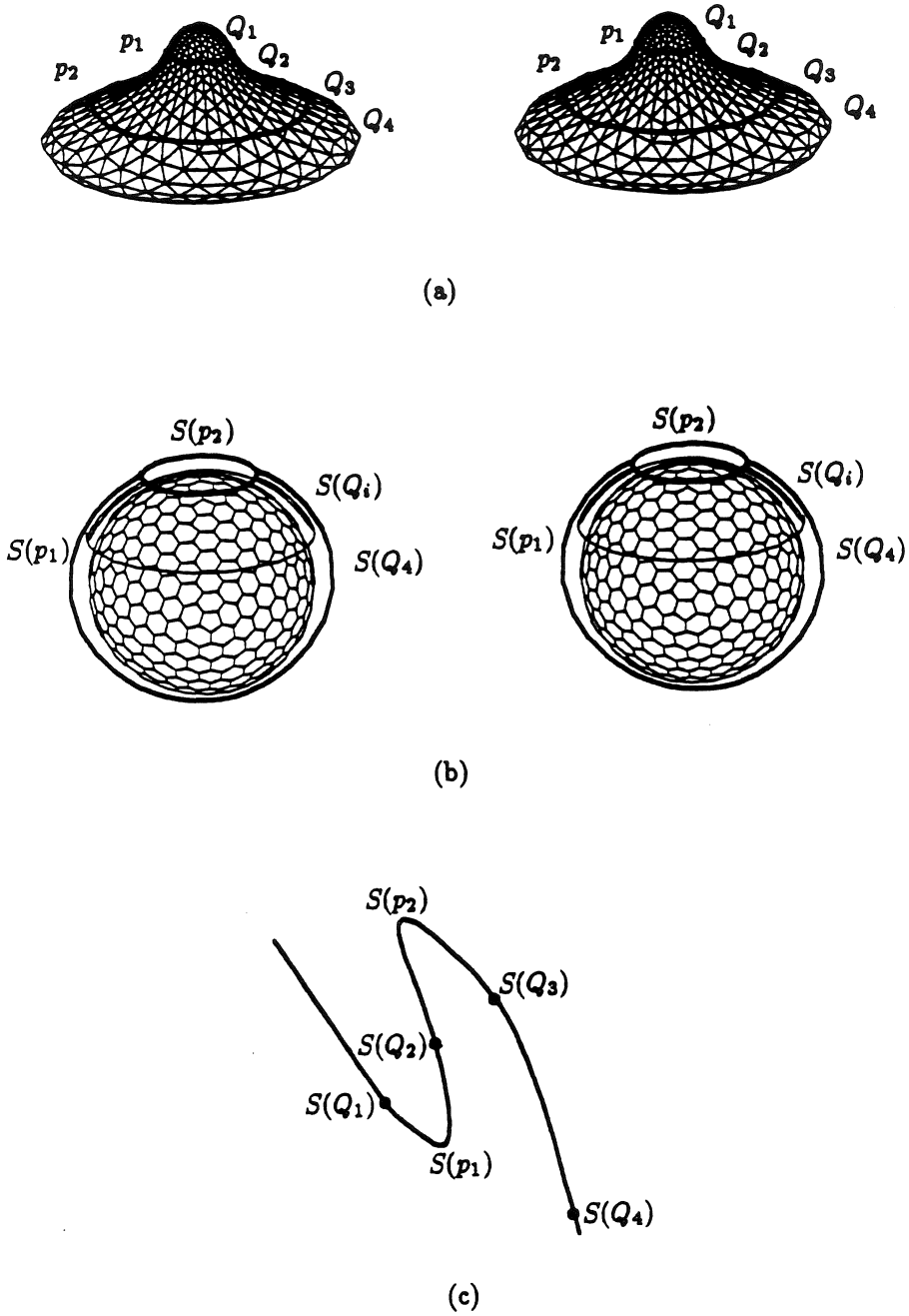


Figure 2.3. Stereoviews of (a) a flying saucer and (b) its spherical image. Curves p_1 and p_2 on the saucer's surface are parabolic curves, which map onto two circles of latitude, $S(p_1)$ and $S(p_2)$, on the unit sphere. Points Q_1, Q_2 and Q_3 have the same surface normal and map onto the same point $S(Q_i)$ on the sphere; Q_4 has a unique normal, and its spherical image $S(Q_4)$ is covered once. (c) shows these regions of the sphere in more detail (from Beusmans, in press).

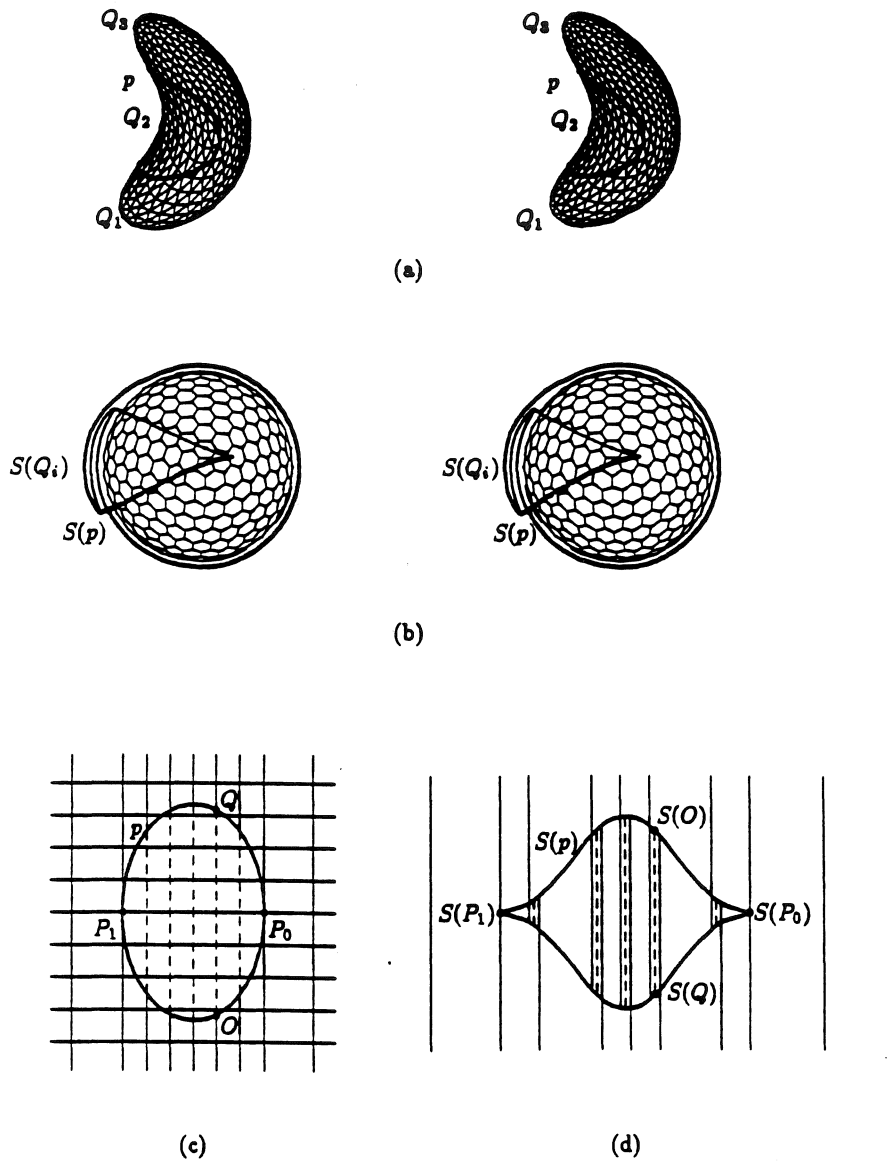


Figure 2.4 Stereoviews of (a) a bean and (b) its spherical image. Points Q_1, Q_2, and Q_3 have the same surface normal and map onto the same point $S(Q_i)$ on the sphere. The spherical image, $S(p)$, of the single parabolic curve p has two cusps only one of which is shown. (c) schematic representation of the two families of lines of curvature in the saddle-shaped region bounded by the parabolic curve p, and (d) its spherical image (only the spherical image of the vertical family of lines of curvature is shown). Dashed lines indicate negative principal curvature. P_0 and P_1 are pedal points: the parabolic curve is tangent to the local direction of zero normal curvature. Each pedal point produces a cusp along $S(p)$, namely at $S(P_0)$ and $S(P_1)$. Note that the region bounded by $S(p)$ is triply covered (from Beusmans, in press).

are perpendicular: Because normal curvature is zero in an asymptotic direction, the surface normal in that direction can only change perpendicular to that direction. Imagine moving along p from O to Q and passing through P_0 (Figure 2.4c). As one gets closer to P_0, the direction of zero principal curvature gets closer to the direction of the parabolic curve itself. causing the angle between the surface normal and the direction of traversal to approach 90° (compare Figures 2.4c and 2.4d). As one passes the pedal point P_0, the direction of zero principal curvature switches to the other side of the parabolic curve causing its spherical image to reverse direction at $S(P_0)$.

Spherical images of objects are of interest in the study of occluding contours because occluding rims map onto great circles on the unit sphere. To be precise, if $\mathbf{V}$ is the viewing direction and we are considering orthographic projection, then the occluding rims map onto the great circle which results from intersecting the unit sphere with the plane perpendicular to $\mathbf{V}$ which passes through the origin of the sphere. In case of perspective projection, the occluding rims map into the zone bounded by a great circle and a circle of latitude related to the size of the object's image (section 4.1 and Figure 4.3; Beusmans, in press). An object's spherical image thus allows for rapidly computing its occluding contours from arbitrary vantage points (Beusmans, in press).

Representing an object by its spherical image and occluding rims by great circles provides intuitive insights into the occurrence of visual events following changes in viewing position. Whenever the great circle crosses the spherical image of a parabolic curve following a change in vantage point, a visual event occurs because either (i) a new occluding rim appears or disappears (lips), or (ii) an existing rim splits into two or two rims join (beak-to-beak). Whenever the great circle crosses the spherical image of a flecnodal curve a swallowtail event occurs (a flecnodal curve consists of flecnodal points, i.e., inflections of asymptotic lines). Scenarios (i) and (ii) are illustrated in Figure 2.5. Spherical images are also helpful in reasoning about points of correspondence on contours (see section 4.1).

If one adds curvature information to the spherical image of an object, one can reconstruct the occluding rim and hence the contour from arbitrary vantage points using the object's spherical image. This has been explored for convex objects by Horn (1984), Little (1983) and Van Hove (1987).

2.3 Occluding contours and equidistance lines

Let S be the smooth surface of a 3D object, I the 2D image of S, and $\rho(S)$ the distance between the observer and points on S. $E_d = \{x \in S | \rho(x) = d\}$ is the set of points on S at a distance d from the observer. Such curves are variously known as equidistance curves, contour lines, or level sets; I will use the term *equidistance curve* to avoid confusion with the term occluding contour, and because I find it the most descriptive of the three.

Figure 2.6 shows equidistance curves on a mannequin; the neighboring curves differ by a fixed distance. The direction of slant $\mathbf{F} = \text{grad} \log \rho_0/\rho$ is everywhere perpendicular to the projection of the equidistance lines, and the magnitude of slant is inversely proportional to their spacing. Thus we note that the spacing becomes infinitesimally small along the mannequin's silhouette, and that the equidistance lines are tangent to the silhouette. Bruce & Giblin (1984, pp. 71, 77) showed that occluding contours are the envelope of the projection of the equidistance lines that intersect the corresponding rims.

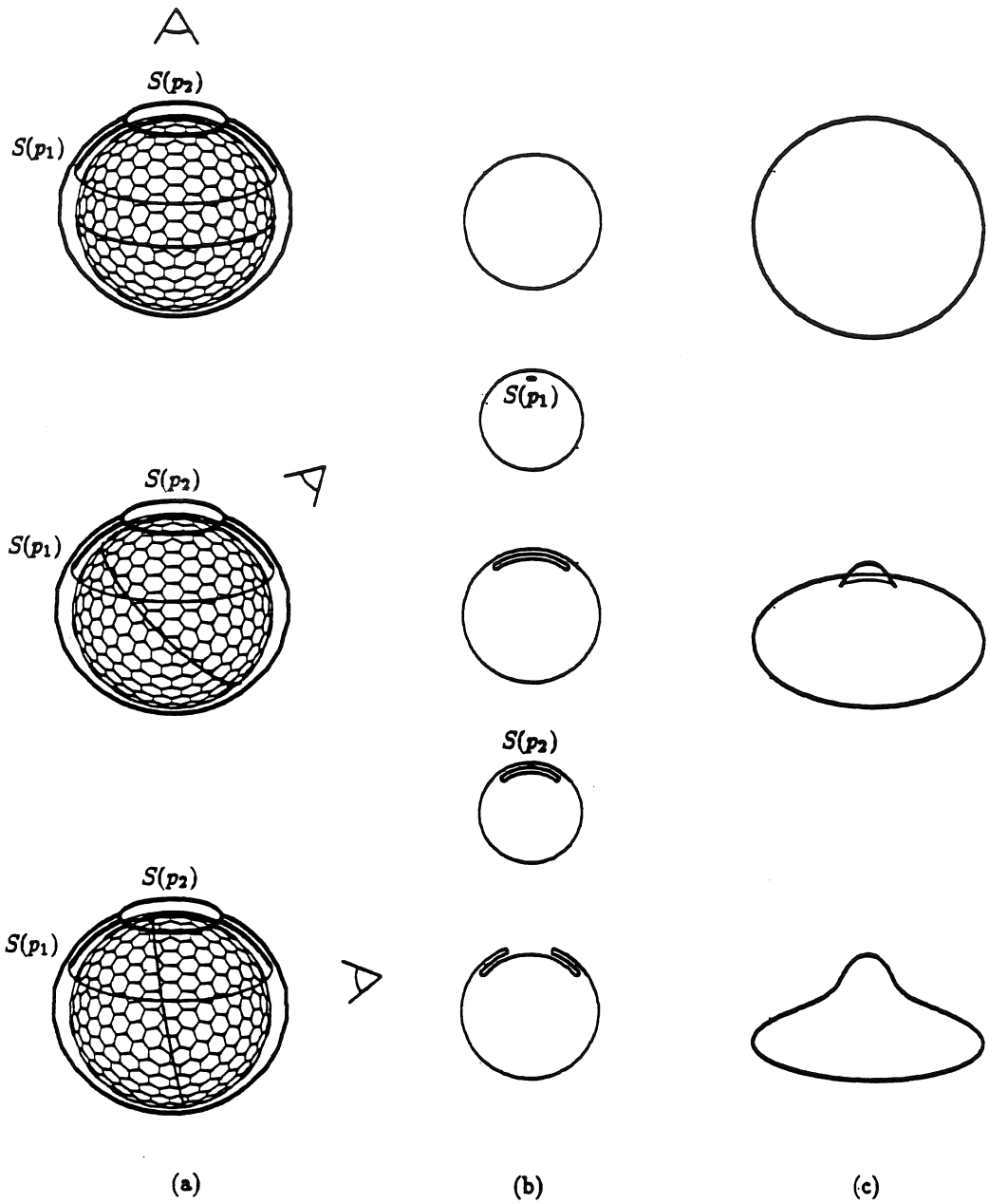


Figure 2.5 Three positions of an observer with respect to the flying saucer: the top row corresponds to a top view, the bottom row to a frontal view, and the middle row to a view in between. (a) shows the spherical image plus the great circle corresponding to the three vantage points. (b) shows the saucer's spherical image along each of the three great circles plus the two transitions that occur when the great circle starts or ceases to intersect the spherical image of a parabolic curve. (c) shows the occluding contours of the saucer (from Beusmans, in press).

The distance function ρ in conjunction with the visual mapping π form an example of mappings of the type $\Re \xleftarrow{f} \Re^2 \xrightarrow{\gamma} \Re^2$, which were analyzed by Dufour (1983). Figure 2.7 lists all the generic forms of a scalar field f in the neighborhood of singularities of γ. In the current application, $f = \rho$ and $\gamma = \pi$. For this special case, the "parapluie transverse" (Figure 2.7, lower left) cannot occur because it features an equidistance line whose image is perpendicular to the occluding contour (indicating that distance along a line of sight is constant, something which is impossible at an occluding contour). Note that distance is maximal at a cusp of the occluding contour, assuming the object is opaque (see Figure 2.7, lower right, and Figure 2.8, bottom row). At a distance extremum along the contour not at a cusp, there is an equidistance line that is tangent to the rim. Because the line of sight is by definition perpendicular to an equidistance line, it follows that the line of sight and the rim are perpendicular. As the two are also conjugated, the rim is in a principal direction at a depth extremum (except for extrema at a cusp). Figure 2.8 shows some examples of equidistance lines and contours for the bean of Figure 2.4.

Along a closed occluding contour, the number of maxima of the distance function equals the number of minima. Therefore, minima and maxima appear and disappear in pairs, say because the vantage point changes or because of object deformation. The maxima at the two cusps of a hump or dent must be separated by at least one minimum. These simple observations together with the fact that the image of equidistance lines is tangent to a contour, and that distance has to increase towards the figure side of a contour allow us to derive qualitative 3D interpretations of contours. Also, if we use contours to represent shape in memory, as in the visual potential discussed in section 2.8, adding the location of distance extrema on the contour and inside the contour, would allow us to reconstruct the 3D shape of an object (upto a relief transformation).

2.4 Motion along occluding contours

Computing motion from smooth curves such as dynamic occluding contours is problematic because there are no identifiable feature points linking the succession of curves (by motion I mean image motion or optical flow produced by a fixed point moving in 3D space). Confronted with this problem, investigators typically assume that the visual system measures local velocity and that the main problem is to integrate these local measurements into a global velocity field. By assuming that local motion detectors treat a small arc of a curve as linear, the "aperture problem" emerges: Motion detectors can only signal the velocity component perpendicular to the curve, the tangential component is lost. To integrate this partial information into a global velocity field, Hildreth (1984) proposed the constraint of smoothness and least variation: Pick the velocity field that is smooth along the curve under consideration and that minimizes $\int |\partial \mathbf{v}(s)/\partial s|^2 ds$, where $\mathbf{v}(s)$ is velocity along the curve.

Although plausible, the constraint of least variation is rather arbitrary. It leads to the correct velocity field only if at every point along the contour the tangent vector $\mathbf{T}$ satisfies $(\mathbf{T} \cdot \partial^2 \mathbf{v}/\partial s^2) = 0$ (Yuille, cited in Hildreth, 1984). This condition holds for translating rigid objects and rigid polyhedra undergoing arbitrary motions. For smooth 3D curves that are moving arbitrarily, the computed velocity field is generally not the correct one. Interestingly, and lending support to the proposed constraint, when humans look at certain curves whose computed velocity fields differs markedly from the true one, they do not perceive

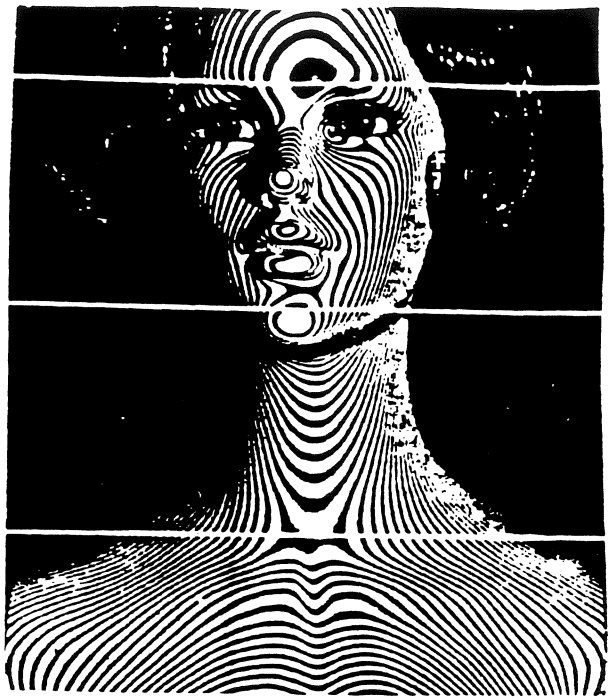


Figure 2.6 Equidistance lines on a mannequin (from Takasaki, 1970). See Koenderink & van Doorn (1978) for a detailed description of this figure and its associated slant field.

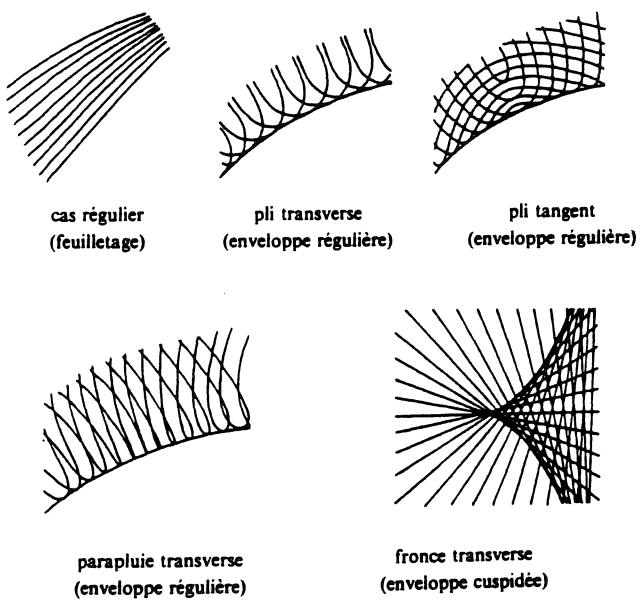


Figure 2.7 Possible forms of the scalar field f in the vicinity of the stable singularities of γ, with $\Re \xleftarrow{f} \Re^2 \xrightarrow{\gamma} \Re$ (from Dufour, 1983).

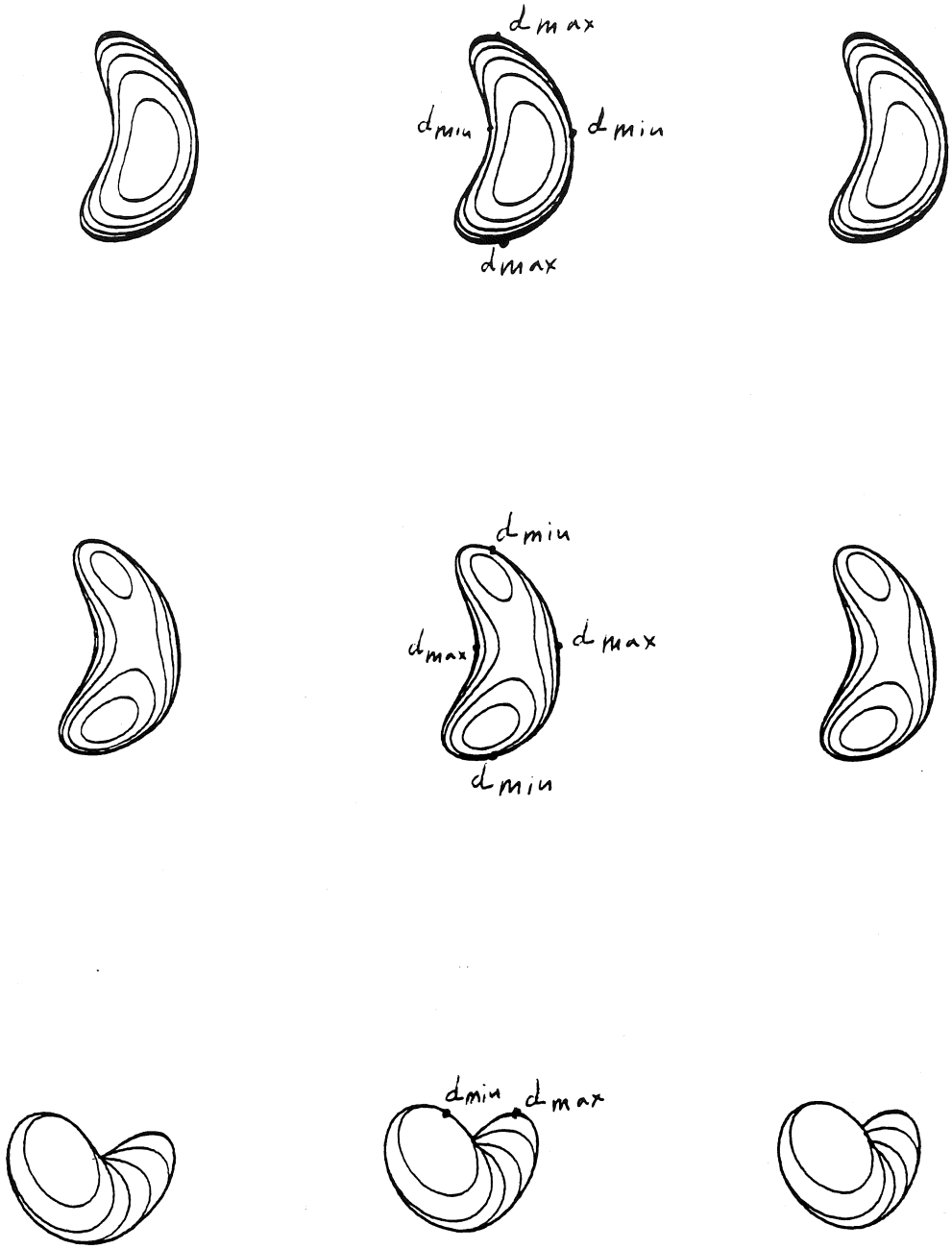


Figure 2.8 Equidistance lines and occluding contours of a bean. Note how the same contour is consistent with different structures of the equidistance lines (compare top and middle rows). The top and middle row differ in their assignment of minima and maxima of distance along the contour: a maximum (minimum) in the top row corresponds to a minimum (maximum) in the middle row.

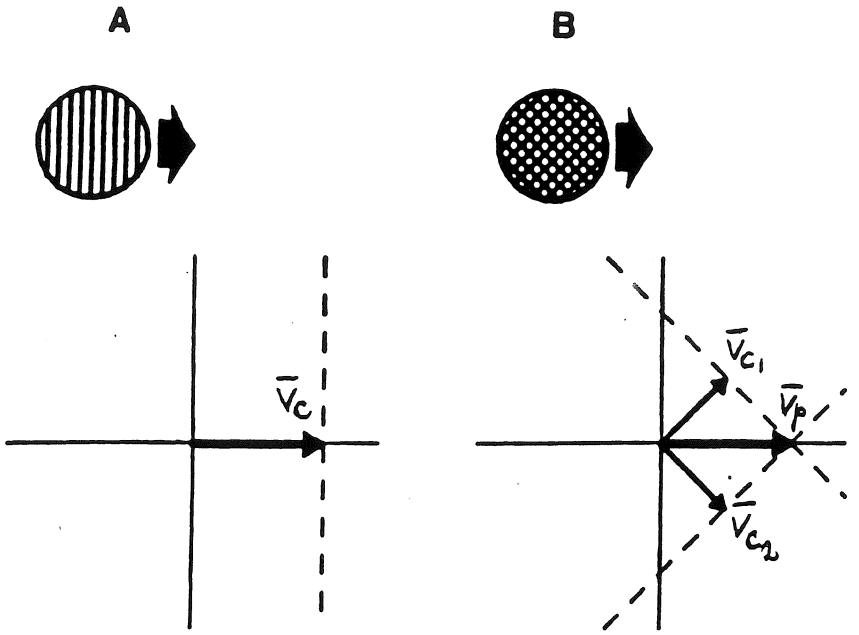


Figure 2.9 (A) A single grating and (B) a 90° plaid consisting of two superimposed gratings whose orientations differ by 90°. The single grating and plaid are both viewed through a circular aperture. The diagrams below the stimuli represent velocity space in which velocities are expressed as vectors; the horizontal (vertical) represents motion in the horizontal (vertical) direction. The grating and the plaid move to the right at the same velocity (represented by the bold, horizontal vector in velocity space). A dashed lines in velocity space specifies all motions that are consistent with the observed motion of a particular grating ($\mathbf{v}_c$ in A, and $\mathbf{v}_{c_1}$ and $\mathbf{v}_{c_2}$ in B). Intersecting the constraints generated by $\mathbf{v}_{c_1}$ and $\mathbf{v}_{c_2}$ yields the pattern motion $\mathbf{v}_p$ of the plaid (adapted from Movshon, Adelson, Gizzi, & Newsome, 1985).

the physically correct state of affairs. For example, an ellipse rotating in the image plane is perceived as deforming (Hildreth's example of the barber pole is misleading; see section 4.1). On the other hand, humans perceive as nonrigid curves such as ojives that are translating, curves for which the constraint of least velocity variation does give the physically correct interpretation (Nakayama & Silverman, 1988).

Strong support for motion detection mechanisms that are selective for orientation comes from studies on the coherence of plaid patterns. A plaid consists of two superimposed, moving sinusoidal gratings that differ in orientation (Figure 2.9). If their contrast and spatial frequency are similar, the two gratings give rise to a percept of coherent motion. The motion of the individual gratings is called *component motion,* that of the coherently moving plaid is called *pattern motion* (see diagrams of velocity space in Figure 2.9). If pattern motion is based upon mechanisms selective for component motion only, then the percept of coherent pattern motion should be disturbed if component motion is disturbed. Adelson & Movshon (1982) tested this by adding dynamic 1-D noise to a plaid. This type of noise looks like randomly moving, parallel stripes of various widths. They found that coherence decreases

most if noise is parallel to either of the two component gratings. This suggests that the perception of coherent motion depends on an orientation-selective stage. Movshon, Adelson, Gizzi, & Newsome (1985) obtained neurophysiological evidence that, in macaque monkeys, the orientation-selective stage occurs in visual area V1 and that pattern motion is computed in MT.

Waxman & Wohn (1987) also started with the assumption that only the velocity component perpendicular to a moving curve is detectable. But rather than making some ad hoc assumptions about the global velocity field, they proceeded by viewing the velocity field along a curve as sampling a 2D velocity field. In other words, they assumed that an image curve is part of a moving surface that would have induced an optic flow field if it had had enough texture. Assuming that this optic flow field is smooth in the neighborhood under consideration, it can be approximated by a second-order polynomial in the image coordinates. This approximation is valid globally if the velocity field derives from a planar surface, and it is valid locally otherwise. The problem is then to recover the twelve coefficients of the two second-order polynomials (one for each component of the velocity vector). As each point of the contour adds one linear equation in terms of these coefficients, we need to sample at least twelve points. Interestingly, bi-quadratic curves such as ellipses are too regular to adequately sample the flow field and hence cannot be used alone to determine all twelve coefficients. The authors referred to this as the "aperture problem in the large." To solve it one needs at least two different bi-quadratic or bi-cubic contours, or a single bi-quartic contour. Alternatively, one can use a single bi-quadratic contour at two successive instances.

The work of Waxman & Wohn shows that it is possible to arrive at a reasonable velocity field and hence a structural interpretation of a moving contour without making ad hoc assumptions such as the constraint of least variation. It is of course an entirely different question whether this analysis has any bearing on human perception. Consider, for example, an ellipse rotating in the image plane. This looks like a deforming figure to human observers. According to Waxman & Wohn's analysis, three views would suffice to get the velocity field consistent with a rotating rigid plane. Does this imply that humans only use two successive views, and cannot integrate information across three views? This would agree with Todd & Brennan's (1990) finding that the ability to discriminate between rigid and nonrigid structures in motion does not improve with more than two views; nor does the ability to indicate which of two lines is the longer one in 3D. It would be interesting to determine exactly how two views constrain the velocity field for a rotating ellipse, and to relate this to a subject's percept.

The visual stimuli used by Hildreth, Adelson & Movshon, and Waxman & Wohn are to some extent special in that their luminance is essentially one-dimensional: The luminance of a grating varies only perpendicular to the grating. For a curve it is even worse; not only does luminance only vary perpendicular to it, luminance also drops to some background value almost instantaneously. In other words, these stimuli are highly degenerate. Reichardt, Schlögl, & Egelhaaf (1988) showed that image motion can be measured locally if only the local intensity distribution is rich enough (so far only the case of translation in the image plane has been studied). Only for certain nongeneric intensity distributions, such as the 1D luminance patterns just described, does the aperture problem emerge. The image intensity associated with an ellipse rotating in the image plane is highly degenerate and might be

the reason its motion is misperceived by human observers. It would be interesting to see if the perception of nonrigidity persists if the ellipse is rendered by a thick line that smoothly varies in intensity along its length.

Faugeras (1990) analyzed the general question whether it is, in principle, possible to infer the structure and motion of a 3D space curve from its projection. If the curve can bend while preserving arc-length, the optic flow cannot be recovered (where by optic flow is meant the projection of the velocity vector of a fixed point on the curve). If the curve is rigid, then its 3D structure and motion can be recovered directly, that is, without having to compute optic flow. This result, however, does not apply to rims as they are not rigid 3D space curves.

Summarizing, Faugeras' (1990) results tell us that, generally, we cannot recover the 3D structure and motion of a sequence of rims from their projection, because they are not the projection of a moving rigid 3D space curve (rims do not even preserve arc length). On the other hand, it is possible to compute the rim's 3D structure if the relative motion of observer and object is known a priori (section 2.7). Thus, the question is whether it is possible to compute this relative motion from the occluding contours themselves without having to simultaneously compute their 3D structure. This is possible for the special case of an object rotating about a frontoparallel axis (sections 2.7 and 4.1); in that case, the axis of rotation and the object's angular velocity can be computed from its dynamic occluding contours. It is not known yet whether the general case of rotation about an arbitrary axis can be solved or not. In section 4.1, I explore one approach and argue that it does not lead to a solution.

2.5 Occluding contours and egomotion

Most animals, including humans, have to move through their environment to look for food or mates, to escape predators or adverse weather conditions, or just for fun. In each case it is obviously important that the animal maintains a clear idea of its direction of motion. This is not a trivial problem if it is to be done on the basis of visual information. Imagine walking over a field in a straight path. While walking along you can look forward in the direction of heading, or to the right to a tree, or to your feet, or you can track a bird across the sky. Yet you never think that you change your heading even though you continuously change the direction in which you are looking and with it the optic flow field. Somehow your visual system extracts the invariant direction of heading.

Egomotion has been the focus of numerous studies ever since Gibson's original work in the early fifties (Gibson, Olum, & Rosenblatt, 1955). However, it was only recently that psychophysicists were able to demonstrate how well human subjects can discriminate between the direction of heading and some arbitrary direction in the image. Warren & Hannon (1990) found thresholds of 1 to 2 degrees for various motions and simulated 3D environments; most importantly, they found that smooth pursuit eye movements did not affect discrimination thresholds. Earlier studies had almost unanimously concluded that humans are actually quite poor at determining heading (for review see Warren, Morris, & Kalish, 1988).

A number of schemes have been proposed to compute the direction of heading. Warren & Hannon (1990) discussed these proposals and tested them empirically. Their findings were consistent with theories that rely on differential motion in the image caused by variations in depth, in particular, along occluding contours. These theories are based on the following

idea, first formulated by Longuet-Higgins & Prazdny (1980). Instantaneously, the motion of an observer can be decomposed into translation and rotation about an axis through its eye. The rotational component of motion induces a retinal velocity field $\mathbf{v}_r$ that is independent of the 3D structure of the environment: it is simply a re-labeling of the visual directions. The translational component of motion, however, induces a velocity field $\mathbf{v}_t$ that does depend on depth: the farther away an object the slower it moves in the image. Also note that translation results in a radial velocity field whose center, the focus of radial outflow (FRO), corresponds to the direction of heading; all velocity vectors point to this center although their magnitude depends on the local depth. The problem, then, is to compute the FRO from the image velocity field $\mathbf{v} = \mathbf{v}_t + \mathbf{v}_r$, that is, in the presence of rotational motion. One solution is to compute local velocity differences; these difference vectors tend to point to the FRO. To see this, imagine what happens at an occluding contour in the image. Pick a point C on the side of the contour that is closer to the observer, and one on the other side, call it F (for close and far). Because C is closer than F, $\mathbf{v}_t(C) > \mathbf{v}_t(F)$, and both point to the FRO; because C and F are infinitesimally close, $\mathbf{v}_r(C) = \mathbf{v}_r(F)$. Therefore, $\mathbf{v}(C) - \mathbf{v}(F)$ points towards FRO. This is not exactly true for image points that are separated by some finite distance, but if one combines velocity differences from the entire visual field the FRO can still be found reliably (Rieger & Lawton, 1985).

2.6 Occlusion in depth and shape perception

An observer moving through an environment with textured, opaque objects at various locations will have visual images in which image detail appears and disappears at occluding contours and in which the image velocity field is discontinuous across occluding contours. Gibson, Kaplan, Reynolds, & Wheeler (1969) pointed out that the appearance and disappearance of image detail (so-called accretion and deletion) is a powerful cue that could enable an observer to infer the relative location of objects in his environment. Kaplan (1969) was the first to simulate and study this aspect of an occluding contour in a laboratory setting. His stimuli consisted of two textured half-fields moving at different velocities (Figure 2.10). He found that accretion and deletion along a straight vertical edge gave the impression of two planes separated in depth, where the half-field with the largest accretion or deletion was seen as behind. Patterns of accretion and deletion at occluding contours are used in computer vision to locate object boundaries and to determine which of the two surfaces at the boundary is the occluder (Mutch & Thompson, 1985).

Kaplan (1969) also found that the relative velocities of the two half-fields could not predict the perceived depth ordering to a significant degree. This agrees with experiments by van Doorn & Koenderink (1982, 1983) on the detectability of discontinuities in velocity fields. Using displays like Kaplan's but with a stationary edge separating the half-fields, they added dynamic white noise to quantify the detectability of various velocity discontinuities. With only a little noise added, the two half-fields appear to be separated by a sharp boundary and the subject has an impression of an ordering in depth. As more noise is added, the boundary and impression of depth disappear whereas the discontinuity in velocity remains detectable. To the extent that accretion and deletion are now undetectable, we can conclude that depth ordering is based upon accretion and deletion and not on velocity discontinuities. It would be interesting to perform an experiment that relates the detectability of accretion and deletion with the percept of depth ordering. Note though that van Doorn & Koenderink's displays are

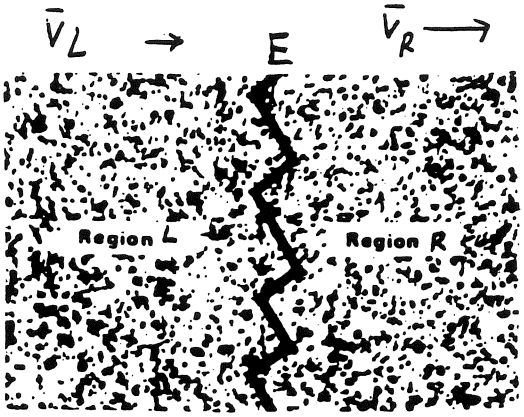


Figure 2.10 Two textured half-fields moving with different velocities v_l and v_r. Texture elements appear and disappear along E, the edge separating the half-fields. The plane associated with the half-field whose texture elements are appearing or disappearing fastest appeared to be behind the plane associated with the other half-field (from Kaplan, 1969).

unecological in that the contour is stationary; the next set of experiments show the motion of the contour relative to that of the half-fields is very important.

It was pointed out recently that Kaplan's displays actually contain another cue for depth ordering. Thompson, Mutch, & Valdis (1985; Thompson, 1989) showed that the image motion of an occluding contour equals that of the occluding surface in the neighborhood of the contour (see also section 4.2 where I show that the velocity field along the occluding contours of a solid rotating about a frontoparallel axis equals that of the corresponding rims rotating about the same axis at the same angular velocity). They developed a scene analysis program which exploits this property of occluding contours. Note that this property is actually more reliable than accretion and deletion because for *curved* objects there will be accretion and deletion on both sides of the contour (as the observer moves his vantage point, the rim on the occluding object changes, causing texture of the *occluding* object to appear or disappear). Thus, simply looking for texture elements that appear or disappear and assigning them to the occluded surface as in Mutch & Thompson (1985) is problematic. On the other hand, it is true that the major portion of accretion and deletion refers to the occluded surface.

Returning to Kaplan's displays, it may be that the percept of depth ordering is not due to accretion and deletion or to differential motion, but to the fact that the contour moved with one of the half-fields, which was therefore seen as in front. Using displays like the ones shown in Figure 2.11 in which a contour is not defined by accretion and deletion but by a single vertical line or by end-stopped horizontal lines, Yonas, Craton, & Thompson (1987) found that human observers indeed perceive as closer the surface that is moving in tandem with the contour. Craton & Yonas (1990) extended these studies and showed that differential motion between contour and occluded surface is more salient than common motion of contour and occluding surface.

The visual stimuli of Craton & Yonas (1990) are nongeneric because the two half-fields

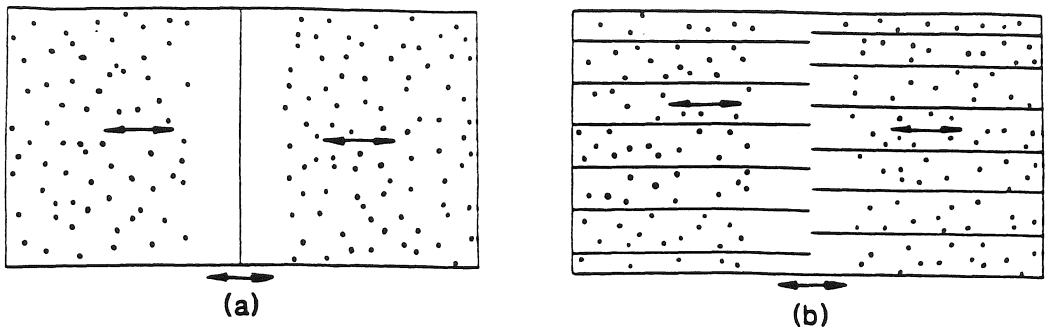


Figure 2.11 Boundary-flow displays with (a) an objective contour and (b) a subjective contour. The "gap" between the central edge eliminates accretion and deletion of texture. The two half-fields move outwards and inwards in phase, and the edge moves in tandem with either the right or the left half-field (from Craton & Yonas, 1990).

were always moving in opposite directions. For observers moving through their environment, motion on either side of a contour will be in the same direction (see section 2.5 on contours and egomotion). If objects can move, then the above situation can arise, although our visual system might track moving objects thereby making image velocity on one side of the contour zero (Thompson, 1989). It would be interesting to repeat these experiments with half-fields moving in the same direction but with different velocities.

The ability of occlusion to order elements in depth is also evident in moving dot displays. If the moving dots simulate a rotating object, the direction of rotation is ambiguous under parallel projection. Occlusion reduces this ambiguity and also leads to implicit occluding contours (Braunstein, Andersen, & Riefer, 1982; Proffitt, Bertenthal, & Roberts, 1984). Occlusion does not affect the shape or amount of depth perceived, it merely disambiguates the percept (Andersen & Braunstein, 1983; Proffitt, Bertenthal, & Roberts, 1984).

2.7 Shape from occluding contours

From static contours we can only gain qualitative shape information. Assuming a figure/ground assignment and that normals point into the figure side of a contour, the sign of curvature of a contour equals the sign of Gaussian curvature along the corresponding occluding rims (Koenderink & van Doorn, 1982). Thus, a convex arc of a contour signals a convex surface patch (concave patches are invisible in opaque objects), and a concave arc signals a saddle-shaped patch. Inflection points separating convex and concave arcs signal parabolic points. In fact, Gaussian curvature $K = \kappa_t \kappa_r$, where κ_t is the curvature of the contour, called *transverse curvature*, and κ_r is the normal curvature of the surface in the viewing direction, called *radial curvature* (Brady, Ponce, & Yuille, 1985; Koenderink, 1984b). Although we can measure κ_t in the image, we only know the sign of κ_r. Hence, we can only determine the sign of K, i.e., recover qualitative shape information.

Giblin & Weiss (1987; Giblin & Soares, 1988) analyzed the special case of object rotation about a frontoparallel axis. In this case, radial curvature, and with it Gaussian and mean curvature, along the rim can be determined exactly. In addition, the 3D location of the occluding rim can be computed. (See section 4.2 for a detailed discussion.) Blake & Cipolla

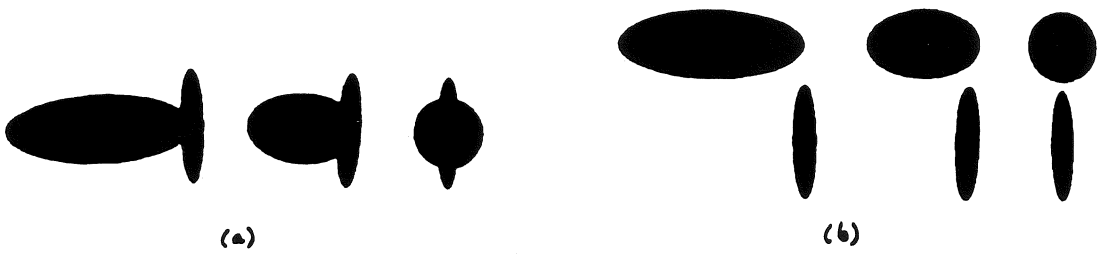


Figure 2.12 (a) Three views of a blimp as it rotates about the vertical. (b) Three views of the two component ellipsoids of the ellipsoid as they rotate about the vertical (from Todd, 1985).

(1989) generalized these results to include arbitrary, but known, motion and perspective projection. Computer simulations showed that surface curvature estimates are highly sensitive to errors in viewer motion. Therefore, Blake & Cipolla proposed that the more robust difference in radial curvature at two locations, or the even more robust ratios of radial curvature differences, be computed.

The role of occluding contours in solid shape perception is still largely unknown. So far it has only been demonstrated that silhouettes of rotating objects, such as in Figure 2.12, can give an impression of a solid, rotating object (Pollick, 1989; Todd, 1985; Wallach & O'Connell, 1953). No systematic investigation into the perceptually necessary or sufficient conditions for this phenomenon has yet been attempted. Nor has anybody tried to measure the 3D shape that is perceived.

The interaction of occluding contours with other sources of shape information also remains to be explored. Ramachandran, Cobb, & Rogers-Ramachandran (1988) noted that the shape of the aperture through which a motion display is viewed markedly influences its 3D interpretation. Specifically, viewing a motion display simulating a rotating cylinder through a conical aperture leads to a percept of a cone instead of a cylinder. Ikeuchi & Horn (1981) used occluding contours to provide boundary conditions in their shape-from-shading algorithm.

2.8 Occluding contours and object recognition

It is not my intention to discuss the general problem of object recognition; I merely want to point out that in one proposed approach to this problem, one that has received considerable attention recently, occluding contours play an important role.

The idea, first formulated by Koenderink & van Doorn (1979), is to represent an object by a collection of all its qualitatively different views. Assuming the shape of an object from a particular vantage point is described by the slant field $\mathbf{F} = \text{grad} \log \rho_0/\rho$, then, qualitatively, its shape is specified by the singularities of the slant field and their spatial ordering. The slant field is singular along occluding contours (slant is infinite) and at minima of the distance function (slant is zero). Figure 2.13a describes one possible view of a torus in terms of singularities and other features of the slant field. Such a qualitative description is called an

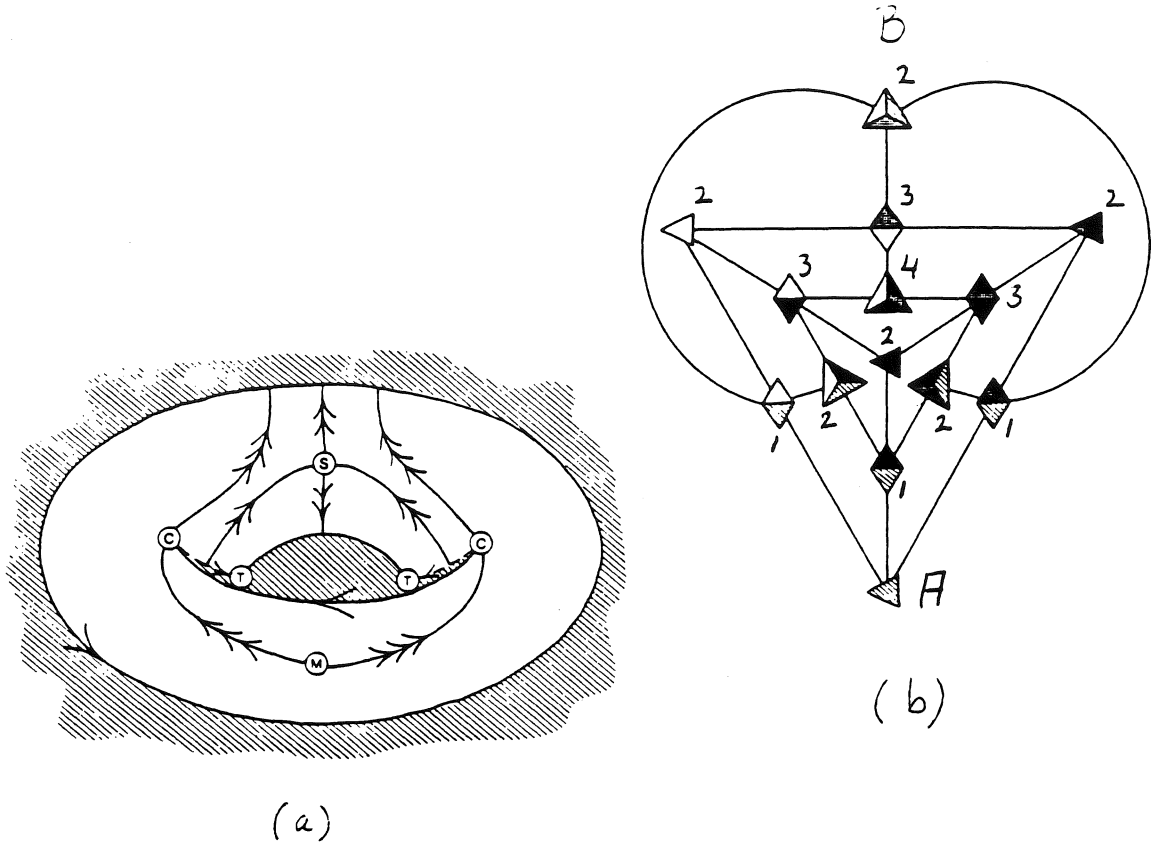


Figure 2.13 (a) A possible view of a torus featuring a figure/ground assignment, occluding contours (single arrow), cusps (C), T-junctions (T), the minimum of the distance function (M), a saddle-point (S), and various lines of steepest increase or decrease in distance (double and triple arrows). (b) The visual potential of a tetrahedron; numbers indicate minimal distance from aspect A (adapted from Koenderink & van Doorn, 1979).

aspect. The aspect of an object is generally stable under small changes in vantage point; visual events, however, change an object's aspect. The structured collection of all aspects of an object is called the *visual potential* of that object; it is a graph whose vertices represent all possible aspects and whose edges specify the spatial relationships between them. Thus, the visual potential of a sphere consists of but one aspect. Figure 2.13b shows the visual potential of a slightly more complex object, a pyramid; it consists of 14 aspects which fall into three classes, namely aspects with one, two, or three visible faces.

The visual potential allowed Koenderink & van Doorn to quantify the notion of complexity of shape. The distance between two aspects, A and B, is simply the smallest number of edges in the visual potential one has to traverse in going from A to B. For example, the distance between aspect A and B in Figure 2.13b is 2. The numbers alongside each aspect indicate their distance from aspect A. This measure of distance, call it d, is a true metric since $d(A,B) = 0$ iff $A = B$, $d(A,B) = d(B,A)$ and $d(A,B) + d(B,C) \geq d(A,C)$. The complexity of a shape is just the largest distance between any two aspects in its visual potential. The intuitively simplest shape, the sphere or more generally the ovoid, has complexity zero,

because its visual potential has only one aspect. The complexity of the tetrahedron of Figure 2.13b is 4.

Plantinga & Dyer (1987) generalized the visual potential by including views from all vantage points, even views that are not separated by a visual event. In particular, they propose that the appearance of objects under orthographic projection be represented in a 4-D space which is the cartesian product of the 2-D image space and the 2-D space of all viewing directions. This so-called aspect space was introduced as a means for computing the visual potential of polyhedra. The original aspect representation explicitly represented the visibility of all faces of a polyhedron. Seales & Dyer (1990) developed a special version of the aspect representation, the so-called rim appearance representation, which is limited to those edges of an opaque polyhedron that can be part of a rim; especially interesting is their explicit representation of T-junctions. Korn & Dyer (1987) describe algorithms to efficiently manipulate and search through a collection of 2-D views. Several computer programs have been written to construct the visual potential of objects, in particular polyhedra (Callahan & Weiss, 1985; Gualtieri, Baugher, & Werman, 1989; Gigus & Malik, 1990; Stewman & Bowyer, 1985); aspect graphs for curved objects are starting to be developed (Ponce & Kriegman, 1990).

The key problem that remains to be addressed for the visual potential is to somehow minimize the potentially enormous number of aspects for even moderately complex objects. One could try to combine the notions of visual potential and scale space (Rieger, 1987), or try to modularize the visual potential (Koenderink, 1984a). Another, more pragmatic approach, would be to limit the number of allowed vantage points (Rosenfeld, 1986).

To determine the relevance of the visual potential for human recognition, Perrett & Harries (1988) had subjects inspect 3D objects (tetrahedra and potatoes) from different vantage points. The objects were positioned at eye level on a turntable which could be rotated by the subjects; objects were not viewed directly, but via a closed-circuit TV in order to eliminate stereo and motion parallax. The subjects were told to inspect a sequence of objects, each for a certain period of time, in preparation for a recognition task. Perrett & Harries measured the inspection time as a function of view. They found that not all views receive equal attention; there are typically a handful of views, the so-called preferred views, that receive most of the attention. This means that there are aspects that receive hardly any attention; furthermore, it was not clear, i.e., in terms of the visual potential model, what made certain aspects preferable. The preferred views did seem to be special in that they often were views of minimal or maximal width.

Although not directly testing the visual potential model, Tarr & Pinker (1989) obtained psychophysical evidence that 2D letterlike characters without a clear intrinsic axis are represented at a number of different orientations. They showed subjects 2D figures at a certain, limited number of orientations in the image, and trained them until they could name them equally fast regardless of orientation. This suggests that these 2D figures are represented at multiple orientations. If a figure is now shown at an unfamiliar orientation, naming time increases in proportion with angular deviation from one of the familiar orientations.

Chapter 3
Depicting occluding contours

In natural images, many things happen simultaneously at occluding contours. Light intensity and texture are typically discontinuous; depth and optic flow are always discontinuous; texture elements appear and disappear; the visual field is segregated into figure and ground; etc. All this gives rise to a problem all too familiar in visual psychophysics. To study the role of a particular cue in shape perception, one has to choose between unrealistic stimuli that have only the desired cue or realistic stimuli that have extraneous shape cues. In fact, it is even worse than that because it is impossible to have only the desired cue without simultaneously making other cues contradictory. In this section, I will discuss various ways of rendering occluding contours.

3.1 Contours rendered as silhouettes

In the few psychophysical studies on the perception of occluding contours, researchers used silhouettes of solids: in their displays, the image is divided into a figure and a ground region, either by making the figure region lighter than the background (Todd, 1985, used a blue figure on a dark background) or darker (Wallach & O'Connell, 1953), or by having a background of stationary dots that become visible or invisible as an opaque object moves in the foreground (Andersen & Cortese, 1989; Pollick, 1989). I find the latter method somewhat problematic for three reasons. First, the contour is defined implicitly and its strength depends on the local dot density. Thus, one does not know what contour the visual system actually infers (of course this applies to any stimulus, but only more so to this one). Second, the figure appears as a hole because it is conveyed by virtue of the absence of anything perceptible, making it hard to perceive a surface. Third, because brightness is a depth cue (brighter things appear to be closer), the fact that the background is brighter than the figure indicates that it is closer, which conflicts with the impression of occlusion. Considering the 2D velocity field defined on the image, note that velocity is everywhere zero except along the silhouette; in other words, the velocity field is discontinuous along the silhouette. In spite of these problems, all of the above display types are effective in the sense that they can elicit a 3D percept in human observers.

One fundamental shortcoming of silhouettes, regardless of how they are rendered, is that internal occluding contours and their associated cusps and T-junctions cannot be displayed because the figure region has no internal structure (T-junctions on the silhouette are evident because of a local discontinuity in the shape of the silhouette and the velocity field along the silhouette). On the positive side, though, silhouettes allow us to determine if internal occluding contours are necessary to get a 3D percept. As human observers can perceive solid shape from the silhouettes of ellipsoids (Andersen & Cortese, 1989; Pollick, 1989; Chapter 6, this dissertation), we can conclude that internal contours are not necessary.

3.2 Contours as discontinuities in light intensity

To make it possible to also render internal contours, one can simply drew bright curves on a dark background to specify the location of occluding contours. As with silhouettes, one has a discontinuous image velocity field and a clear separation of figure and ground. The impression of 3D shape elicited by these displays is slightly less than that of silhouettes because figure and ground have the same appearance: in addition, there is the possibility that these curves are interpreted as the projection of a rotating wire frame rather than the occluding contours of a rotating solid.

To reduce the tendency to interpret curves as wire frames instead of bounding contours of a solid object, one might use stimuli that are half way between silhouettes and curves. The idea is to use very wide curves whose intensity increases smoothly going into the figure. Thus there is an abrupt change in luminance only at the occluding contour. This highly stylized luminance gradient corresponds to a situation in which the light source is somewhere behind the observer.

An interesting question is what shading in the neighborhood of occluding contours might tell us about the direction of the light source. One of the big problems in shape from shading is finding the direction of the light source; without that information, image intensities cannot be translated into surface orientations. The approach so far has been to make statistical assumptions about the distribution of normals of the surface being imaged, i.e., the shape of the surface and then to compute the illuminant direction (Lee & Rosenfeld, 1985; Pentland, 1982). Instead of making strong a priori assumptions about surface shape, one could focus on the neighborhood of occluding contours where surface orientation is known to be highly constrained. The idea is to take a global view of the occluding contours of a compact solid object, and to compare, say, image intensities of surface patches whose normal is *known* to be antiparallel. For example, if light intensity is the same everywhere along the contour the light source can only be behind the observer. Complications arise in the form of attached shadows if the light source is not behind the observer, in the form of multiple light sources, ambient light, etc. (Ikeuchi & Horn, 1981, developed a shape from shading algorithm in which occluding contours provide boundary conditions but only locally; they did not take the global view suggested here.) If this idea is valid with regard to human perception, small changes in illumination along occluding contours should have large effects on perceived shape.

3.3 Contours as discontinuities in velocity fields

Discontinuities in a velocity field are hard to dissociate from accretion and deletion boundaries, i.e., boundaries defined by the appearance and disappearance of image elements. Yonas, Craton, & Thompson (1987) and Craton & Yonas (1990) developed visual stimuli in which there is a contour without accretion and deletion (Figure 2.11). However, these stimuli also lack a discontinuity in image velocity.

The familiar dilemma now takes the form of having image motion be realistic and hence introducing an additional cue to shape, or being unrealistic and thus contradictory to what the contour indicates. That the shape of a contour and local image motion can interact in an interesting manner was demonstrated recently by Ramachandran, Cobb, & Rogers-Ramachandran (1988). They found that if moving dots simulating a rotating cylinder are viewed through a triangular aperture, a cone instead of a cylinder is perceived.

3.4 Contours as uncorrelated regions in stereo images

The perception of depth elicited by random dot stereograms is commonly attributed to disparity between correlated regions of the left and right image (Julesz, 1964; Marr, 1982). Logically at least, it is just as plausible to say that it is disparity between uncorrelated regions that leads to a perception of depth. While not addressing this issue directly, Julesz (1964) formulated the rule that "undetermined areas (areas without disparity) are perceived at the depth of the most distant adjacent determined area (area with disparity)." Phenomenally, this did indeed happen in his stereograms in which a rectangle stands out in front of or behind a planar background. But this raises an interesting point: Maybe the uncorrelated areas (or the boundary between an uncorrelated and correlated area) are matched first and then impart depth to the correlated regions they circumscribe, instead of the other way around.

The results of several experiments add credence to this interpretation. Richards (1977) found that the perception of depth in random dot stereograms presented for 200ms to avoid vergence eye movements is much less as compared with stereograms which also have monocular contours. Ramachandran, Madhusudhan Rao, & Vidyasagar (1973) and Ramachandran & Cavanagh (1985) showed that subjective contours and contours defined by a difference in texture can "capture stereo": The regions enclosed by these contours are perceived at the same depth as the contours even though the regions themselves have no disparity. Using ambiguous stereograms consisting of regularly spaced dots, Mitchison & McKee (1985, 1987) showed that perceived depth is determined by disparity at the edges of the stereogram and not by point-to-point matches in between the edges. Gillam & Borsting (1988) found that uncorrelated texture speeds up stereoscopic processing, and they surmised that stereopsis actually starts from uncorrelated regions.

These results suggest that the relatively small (in terms of area in the image) contours and uncorrelated regions determine how the large correlated areas are perceived. Stevens & Brookes (1987a) argue that depth perception is analogous to brightness perception. Just as brightness of an equiluminant patch is largely determined by its contrast with surrounding patches rather than its own light intensity, so is its depth determined by disparity *differences* with surrounding areas rather than by its own *constant* disparity.

Shimojo, Silverman, & Nakayama (1988) obtained evidence for a connection between occlusion and stereopsis. They studied the perception of depth obtained by moving a narrow vertical bar behind an opening and varying its visibility to the left and right eye. The bar is seen as behind the opening if, when moving to the right, it is first seen by the right eye and then by the left eye; perceived depth increases with the delay in visibility by the left eye. If the bar is first seen by the left eye, no depth is perceived. This implies, according to Shimojo et al., that the depth perceived in these displays is not due to conventional binocular stereo mechanisms which would have caused the bar to appear in front of the slit, but to mechanisms that analyze the occlusion relationships between the left and right image.

In view of these indications of the importance of uncorrelated regions in stereopsis and of a connection between occlusion and stereopsis, it would be interesting to construct ambiguous stereograms of compact solid objects and to study how shape information is propagated from uncorrelated regions—i.e., along occluding contours—to the correlated but ambiguous ones.

Chapter 4
Inferring object shape and motion from occluding contours

In this chapter, I will discuss various ways to compute the 3D shape and motion of a rotating solid object from its occluding contours. It is important to realize from the outset that there is no such thing as "*the* shape" of an object, or, as Koenderink (1990, p.15) put it: "things do not 'have a shape' the way Santa Claus has a red suit." Shape is something that arises through the interaction of an observer with an external object and exists, inside the observer, as a model of certain aspects of that object.

To illustrate these remarks, consider the following five models an observer might entertain regarding a solid object (Equation 4.1). In M_1 we simply assign depth to points in the optic array: With each visual direction (θ, ϕ) we associate the scalar ρ indicating the distance from the observer to the nearest point on the object's surface. In M_2 we indicate the orientation of the local tangent plane in the direction (θ, ϕ) in terms of the slant (σ) and tilt angle (τ) with respect to that viewing direction. M_3 simply combines M_1 and M_2. M_4 specifies the local conjugation relationships, $\mathcal{C}$, say in the form of Dupin's indicatrix (Figures C.1, C.2, and C.3 in Appendix C). In M_5, we associate with each visual direction the Gaussian curvature at the corresponding surface point.

$$\begin{aligned} M_1(\theta,\phi) &= \rho, \\ M_2(\theta,\phi) &= (\sigma,\tau), \\ M_3(\theta,\phi) &= (\rho,\sigma,\tau), \\ M_4(\theta,\phi) &= \mathcal{C}, \\ M_5(\theta,\phi) &= G. \end{aligned} \tag{4.1}$$

To compare the different models we need to define a *purpose* $\mathcal{P}$ to induce a metric on the space of models. Two models are equivalent with respect to $\mathcal{P}$ if they can answer the question posed by $\mathcal{P}$. For example, if the current purpose is to guide the observer's motor system to throw something at the object, M_1 and M_3 are equivalent in a positive sense; the remaining models are equivalent in that they cannot guide the motor system.

Now suppose the purpose is to predict which regions of the object will become invisible, say during the next second, if the object were to translate by $\mathbf{T}$ (assume the object is opaque and consider only local occlusions). M_1 can say where a feature point will move to in the optic array but not whether it will remain visible because that requires knowing the local surface orientation; therefore M_2 and M_3 can answer the question. The other two models fail because they cannot predict future surface orientations. However, M_4 can give a partial answer in the sense that it can indicate which surface points will disappear together. This is

possible because the direction of a future rim is locally conjugated to the pre-image of $\mathbf{A}_t$, the projection of $\mathbf{T}$ onto the viewing sphere (Koenderink & van Doorn, 1976): M_4 immediately gives us the direction conjugated to the direction of $\mathbf{A}_t$, which itself can easily be computed from $\mathbf{T}$. This is illustrated in Figure 4.1, which shows the conjugation relations at various locations on a sphere. The sphere translates to the right, parallel to the image plane, by $\mathbf{T}$. The directions conjugate to the pre-image of $\mathbf{A}_t$ are indicated by upwardly pointing arrows. These arrows indicate the local direction of the rim with $\mathbf{A}_t$ as the projection of the corresponding viewing direction.

As a final example, suppose we have to decide whether a surface point is elliptic or not, something which could be useful for a recognition scheme based on a qualitative description of shape. M_4 and M_5 both provide an immediate answer: A point is elliptic if $G > 0$ (M_4), or if $\mathcal{C}$ does not have any asymptotic directions (M_5). The other models cannot answer this question without considering neighboring points.

Summarizing, M_4 and M_5 are equivalent with respect to the purpose of determining convexity. For the purpose of predicting occlusions, M_4 is partially equivalent to M_2 and M_3, which are themselves equivalent. If we have to estimate distance, M_1 and M_3 are equivalent.

Thus there are many different models an observer might entertain regarding the spatial structure of an object. A model might serve more than one purpose. For example, M_4 can determine which regions of an object will become invisible under a certain motion of the object or observer, but it can also serve to subdivide the object's surface into elliptic and hyperbolic patches. Conversely, more than one model might be able to address the same question or purpose. Moreover, if we allow local computations within an existing model, say in the form of visual routines (Ullman, 1984b), the range of questions a model can address can be greatly extended. For example, let the slant field or nearness gradient be $\mathbf{F} = \text{grad}\log(\rho_0/\rho)$, so that the slant angle σ is given by $\sigma = \text{atan}|\mathbf{F}|$ and the tilt angle τ equals the direction of $\mathbf{F}$ in the optic array. Thus it would be easy to go from M_1 to M_2. Or to give another, slightly more plausible example in the context of human visual perception where ρ is not known accurately but $\mathbf{F}$ probably is, given $\mathbf{F}$ in the neighborhood of a point P it is straightforward to construct the conjugation relations. Let Q be a neighboring point of P in the direction dr in the optic array; then the direction perpendicular to $\mathbf{F}_P - \mathbf{F}_Q$ is the image of the direction conjugated to the pre-image of dr (Koenderink & van Doorn, 1978). See section A.7 for a proof and Figure 4.2 for some examples. Perhaps such a visual routine operates on parts of the image that have high resolution.

Clearly, a psychophysicist wishing to determine whether the human visual system uses a particular shape model to perform a certain task is faced with a big challenge. Ideally one would like one model to succeed under certain conditions and all the others to fail. Barring this ideal situation one has to develop detailed *quantitative performance* measures, that will allow one to compare the different models.

In the present chapter, I investigate what information about the shape and motion of an object can be computed from a sequence of its occluding contours, and I formulate several models for the perception of shape. In Chapter 6, I use these models to derive performance measures for a psychophysical discrimination task.

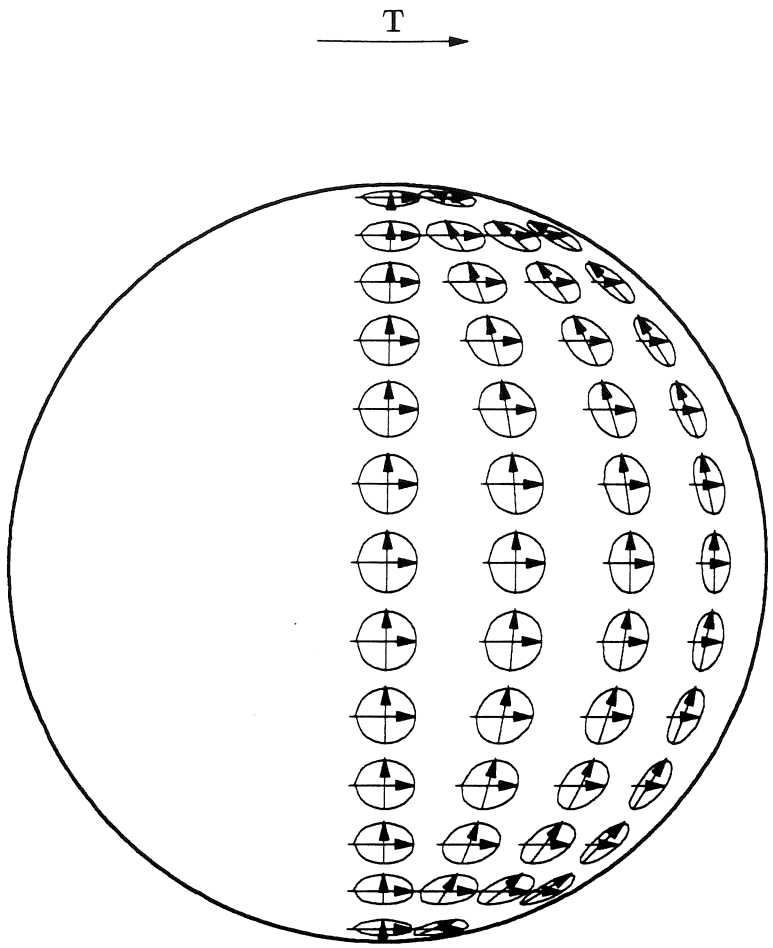
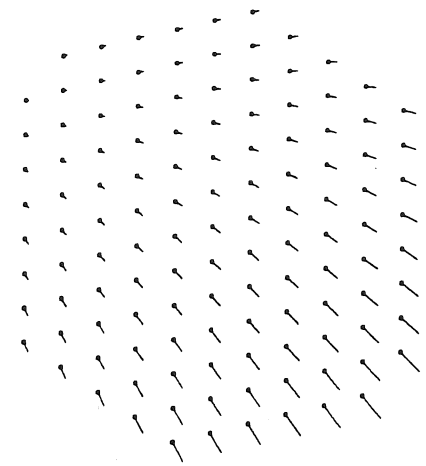
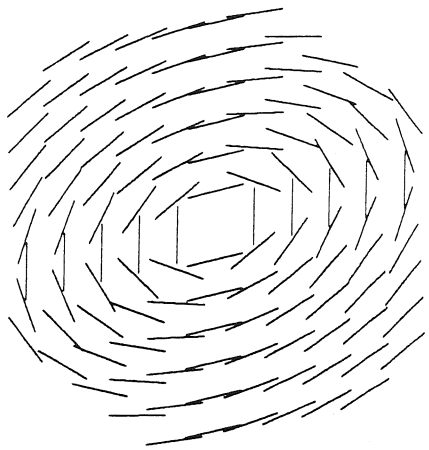


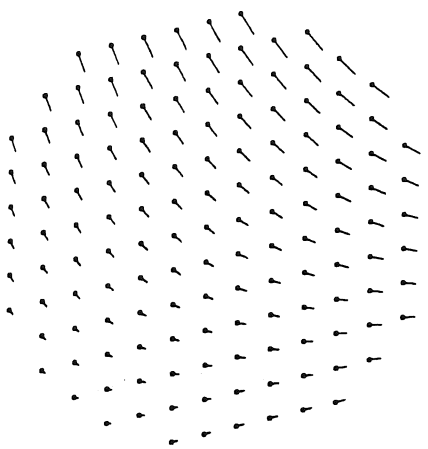
Figure 4.1. The projection of Dupin's indicatrix at various locations on a sphere. **T**, the direction of motion of the sphere relative to the observer, is to the right and parallel to the image plane. The two arrows inside each indicatrix indicate conjugate directions. By tracing the arrows conjugate to $\mathbf{A}_t$, one obtains future rims.



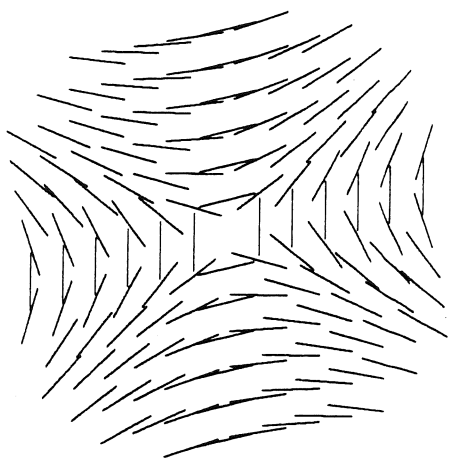
(a)



(b)



(c)



(d)

Figure 4.2. (a) Slant field $\mathbf{F}$ near the origin O of the elliptic paraboloid $z = x^2 + 2y^2$. (b) Directions conjugated to $\overline{OP}$, where P is any of the points in the image. (c) Slant field near the origin of the hyperbolic paraboloid $z = x^2 - 2y^2$. (d) Directions conjugated to $\overline{OP}$; note the two approximately diagonal asymptotic directions. Both paraboloids were first rotated about the vertical by 30° and then rotated about the horizontal by 25°.

The theoretical development in the remainder of this chapter is guided by the following three-stage model for the inference of shape from contours. In the first stage, the motion of the object producing the contours is determined. This information is propagated along the contour to obtain a velocity field, which guides the second stage, the computation of relative depth and 3D shape along the rim, the contour's pre-image. In the third stage, these local shape descriptors are integrated into a global percept of solid shape.

4.1 Computing object motion from occluding contours

My approach will be to try to locate those points on a sequence of occluding contours that are in correspondence and to use them to compute the object's motion.

Definition 4.2. A point P_1 on the contour at time t_1 and a point P_2 on the contour at time t_2 are in *correspondence* iff they are the image of the same point on the object; that is, if $\pi_1^{-1}(P_1) = \pi_2^{-1}(P_2)$, with π_1 and π_2 the projection at time t_1 and t_2. They will be referred to as *points of correspondence (POC)* between these two contours. If a point on the surface of a rotating object is projected onto the contour for more than two views, it will be called a *fixed point.* Clearly, all fixed points are pre-images of points of correspondence but not vice versa.

The Gaussian image of an object is a convenient tool to reason about points of correspondence. Let $N : M \rightarrow S^2$ be the Gauss map of the 2D manifold M with S^2 the unit or Gauss sphere. The symbol N serves as a reminder that the Gauss mapping is a mapping by parallel normals.

Definition 4.3. The *cover* of the Gauss sphere at $s \in S^2$ is the number of points P on the surface for which $N(P) = s$.

I will only consider objects whose Gaussian image has a finite cover everywhere it is defined. This means that I will exclude such pathological objects as infinitely many sheets stacked on top of each other; this could be viewed as a global structure causing the cover of the Gauss sphere to become infinite. There are also local and less pathological structures that can give rise to an infinite cover. Because Gaussian curvature is the determinant of the Gauss map (Hilbert & Cohn-Vossen, 1983), the Gauss map is singular at parabolic points; in other words, the Gauss map is not one-to-one at parabolic points, although it does not imply that it is infinitely-many-to-one. The latter occurs for planes or cones, both of which consist entirely of parabolic points. Thus, I exclude developable surfaces from consideration, which implies that parabolic points are only allowed to occur in curves. Furthermore, I have to exclude parabolic curves that are contained in a plane. For example, the two parabolic curves of a torus each lie in a plane; consequently, the normals along each of them are identical, causing their Gaussian image to consist of only one point whose cover is infinite. Thus, I restrict the class of objects to those whose parabolic curves are only locally planar as at a pedal point (Figure 2.4).

Theorem 4.4. The cover of the Gaussian image of a compact object is odd everywhere except along folds where it is always even.

Proof. The cover of a compact, convex object is everywhere 1. By deforming it as if it were a piece of clay (i.e., without changing its topology), and introducing hyperbolic and concave patches, we create folds in the Gauss map. Across folds the cover always increases

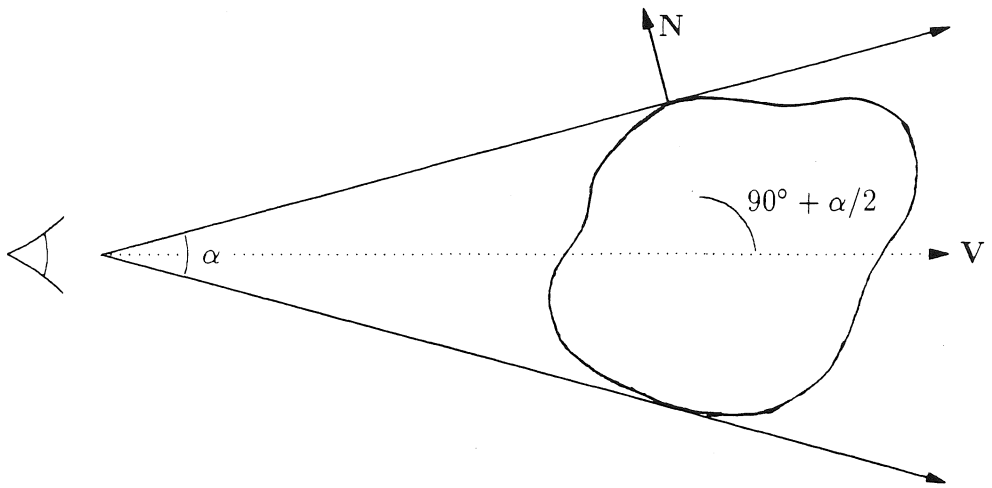


Figure 4.3. Viewing cone with top angle α of an object under perspective projection. For a point to lie on the rim, its surface normal $\mathbf{N}$ has to satisfy $90° \leq \arccos(\mathbf{N} \cdot \mathbf{V}) \leq 90° + \alpha/2$.

or decreases by 2, so that the cover at a point on the Gauss sphere remains odd. Along folds themselves, the cover is even. □

Given a rim R on an object, what is its Gaussian image $N(R)$? Under orthographic projection with viewing direction $\mathbf{V}$, $N(R)$ is a great circle on the Gauss sphere, and the plane containing this great circle is perpendicular to $\mathbf{V}$. Thus, for all $s \in N(R)$, the surface normal at $N^{-1}(s)$ is perpendicular to $\mathbf{V}$. Under perspective projection, the situation is more complicated as each point on the object is seen under its own viewing direction. Let the minimal disc that contains the object's image subtend a visual angle α, and let $\mathbf{V}$ be in the direction of the disc's center; in other words, $\mathbf{V}$ is the axis of the viewing cone with top angle α (Figure 4.3). A point on the object can be part of the rim if its surface normal satisfies $90° \leq \arccos(\mathbf{N} \cdot \mathbf{V}) \leq 90° + \alpha/2$. Consequently, the Gaussian image of potential rim points is contained within the zone on the Gauss sphere bounded by the great circle associated with $\mathbf{V}$ and the circle of latitude at $\alpha/2$ in the direction of the vantage point. For an object with a diameter of .2m viewed from 1m, $\alpha/2 = \text{atan}(.1) \approx 6°$. The farther the vantage point is from the object the smaller is α and the zone of potential rim points; as the vantage point recedes to infinity, the zone increasingly approximates the great circle (also see Figure 4.6). In the sequel, I will only consider orthographic projection.

Now consider two arbitrary views of an object corresponding to viewing directions $\mathbf{V}_1$ and $\mathbf{V}_2$. Let $N(R_1)$ be the Gaussian image of the rim associated with $\mathbf{V}_1$, $N(R_2)$ the one associated with $\mathbf{V}_2$. $N(R_1)$ and $N(R_2)$ are two great circles on the Gauss sphere. Two great circles either completely overlap or intersect in two points. If they overlap, $\mathbf{V}_1 = \mathbf{V}_2$; the two contours are related by a rotation about the viewing direction $\mathbf{V}_1 = \mathbf{V}_2$, and consist entirely of points of correspondence (POCs). If the two great circles intersect in two points, there will be an even number of POCs because each of the two intersections contributes an

odd number of POCs (for compact objects that is). If the object is opaque, some of the POCs might not be visible. Thus, two arbitrary contours of an object have at least two POCs.

If we have a sequence of rims, R_t, arising from an object rotating about some fixed axis, the intersections of their successive Gaussian images, i.e., $N(R_t) \cap N(R_{t+1})$, lie on two circles of latitude on the Gauss sphere (Figure 4.4). In the continuous case where the successive viewpoints are infinitesimally close, these intersections form the envelope of the family of Gaussian images (Bruce & Giblin, 1984, p.80, show that the envelope of a family of curves is the limit of intersections of nearby curves in that family).

Theorem 4.5. If an object rotates about a fixed axis which is slanted by θ_0 with respect to the image plane and if the amount of rotation at time t is $\phi_0(t)$, then the envelope of the family of great circles $N(R_{\phi_0})$ on the Gauss sphere consists of the two circles of latitude $\theta = \theta_0$ and $\theta = 180° - \theta_0$.

Proof. Assume the center of the Gauss sphere is fixed at the origin and assume the standard spherical coordinate system (Figure A.1). Let the current viewing direction be $\mathbf{V} = (90° + \theta_0, \phi_0)$, and R the corresponding rim. $N(R)$ consists of all those points (θ, ϕ) on the Gauss sphere that are perpendicular to $\mathbf{V}$, that is,

$$\begin{aligned} \mathcal{G}: \quad & \sin(90° + \theta_0)\sin\theta(\cos\phi_0\cos\phi + \sin\phi_0\sin\phi) + \cos(90° + \theta_0)\cos\theta \\ = \; & \cos\theta_0\sin\theta(\cos\phi_0\cos\phi + \sin\phi_0\sin\phi) - \sin\theta_0\cos\theta \\ = \; & 0. \end{aligned} \tag{4.6}$$

If the observer rotates about the z-axis (thus changing only the ϕ_0 component of $\mathbf{V}$), the family of great circles can be parametrized by ϕ_0. One can also view this as the object, and with it the Gauss sphere, rotating about the z-axis, while the viewing direction remains fixed. In any event, the axis of rotation is slanted by θ_0 with respect to the image plane. The envelope of the family of great circles on the Gauss sphere is defined by $\mathcal{G} = \partial\mathcal{G}/\partial\phi_0 = 0$:

$$\frac{\partial \mathcal{G}}{\partial \phi_0} = \cos\theta_0\sin\theta(-\sin\phi_0\cos\phi + \cos\phi_0\sin\phi) = 0. \tag{4.7}$$

Equation 4.7 is true if $\theta_0 = 90°, \theta = \{0°, 180°\}$, or $\phi = \{\phi_0, \phi_0 + 180°\}$. Substituting $\theta_0 = 90°$ into Equation 4.6 yields $\theta = 90°$. Similarly, if we substitute $\theta = 0°$ $(180°)$ into Equation 4.6, we get $\theta_0 = 0°(180°)$. Both are special cases of the general solution which we obtain by substituting the two values for ϕ into Equation 4.6. Substituting $\phi = \phi_0$ yields:

$$\mathcal{E}: \quad \cos\theta_0\sin\theta - \sin\theta_0\cos\theta = 0. \tag{4.8}$$

Solving for θ, we get $\theta = \theta_0$. Similarly, substituting $\phi = \phi_0 + 180°$ gives $\theta = 180° - \theta_0$. Thus, the envelope consists of the two circles of latitude $\theta = \theta_0$ and $\theta = 180° - \theta_0$. □

There are two special cases. If $\theta_0 = 90°$, that is, the axis of rotation coincides with the viewing direction, then the rim R_t and its Gaussian image $N(R_t)$ do not change over time. Consequently, the envelope of $N(R_t)$ is $N(R_t)$ itself. The contour does not change shape, it merely rotates in the image.

In the second special case, $\theta_0 = 0°$, that is, the axis of rotation is frontoparallel. The two circles of latitude that make up the envelope now degenerate into two points, namely

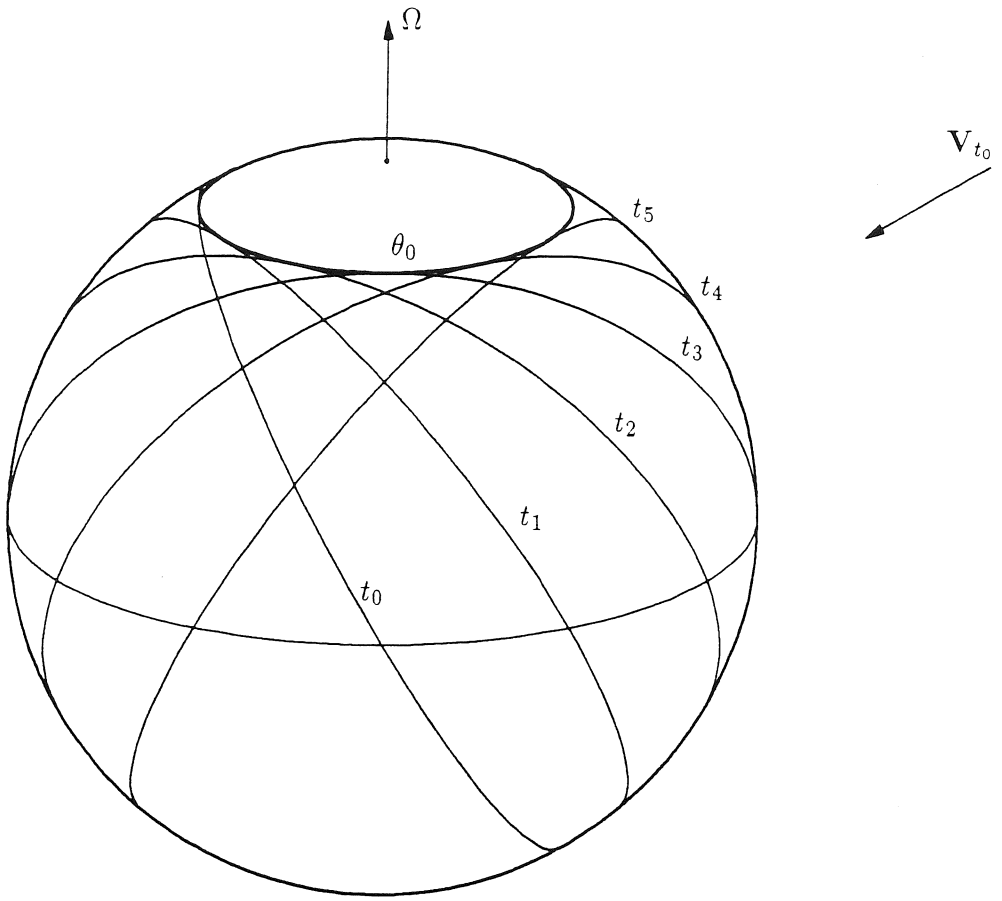


Figure 4.4. Gaussian images of successive rims. Under orthographic projection, the Gaussian image of a rim is a great circle. If the object is rotating about a fixed axis Ω which is slanted θ_0 degrees with respect to the image plane, the Gaussian images of successive rims are all tangent to the circles of latitude $\theta = \theta_0$ and $\theta = 180° - \theta_0$. In other words, $\theta = \theta_0$ and $\theta = 180° - \theta_0$ form the envelope of these Gaussian images ($\theta_0 = 30°$; the angle between the viewing directions $\mathbf{V}_t$ and Ω is 60°). If the current rim is at t_0, the others succeed it by the indicated time; the object rotates 30° between successive views. The Gauss sphere has been slanted towards the viewer to show the circle of latitude $\theta = \theta_0$ more clearly; $\theta = 180° - \theta_0$ is invisible.

the north and south poles of the Gauss sphere. This means that the Gaussian image of the points of correspondence remains the same as the object rotates; in other words, these points of correspondence are images of fixed points. This is not true for the general case $0° < \theta_0 < 90°$: now the points of correspondence change over time. To further clarify this, recall that any two successive views, say v_{i-1} and v_i, of a compact object have an even number of POCs. However, the POCs of views v_{i-1} and v_i need not be those of v_i and v_{i+1}. In terms of great circles on the Gauss sphere, note that three arbitrary great circles generally do not intersect in two points; if they do, they are related by a rotation about an axis through these two points, that is, an axis parallel to the image plane.

An interesting property of the viewing direction, the axis of rotation, and the surface normal at the pre-image of points of correspondence is that they are coplanar (first pointed out to me by F. Pollick in a different context). The viewing direction is $\mathbf{V} = (90° + \theta_0, \phi_0)$ $= (\cos\theta_0 \cos\phi_0, \cos\theta_0 \sin\phi_0, -\sin\theta_0)$, the axis of rotation is $\mathbf{\Omega} = (0, 0, 1)$, and the Gaussian image (i.e., the normal) of the pre-image of the points of correspondence p associated with $\mathbf{V}$ is

$$\mathbf{N}(\pi^{-1}(p)) = (s_{\theta_0} c_{\phi_0}, s_{\theta_0} s_{\phi_0}, c_{\theta_0}). \tag{4.9}$$

or $180° - \theta_0$ instead of θ_0. In either case,

$$\begin{aligned} [\mathbf{V}\mathbf{\Omega}\mathbf{N}] &= (c_{\theta_0} s_{\phi_0}, -c_{\theta_0} c_{\phi_0}, 0) \cdot (s_{\theta_0} c_{\phi_0}, s_{\theta_0} s_{\phi_0}, c_{\theta_0}). \\ &= 0. \end{aligned} \tag{4.10}$$

This property can be used to compute the points of correspondence if one knows the axis of rotation and the viewing direction.

Figure 4.5 shows the POCs of ellipsoids rotating about an axis which is slanted by various degrees and which goes through the center of the ellipsoids. Note that even for such a simple object as an ellipsoid the POCs form very complicated curves; the sole exception is zero slant, that is, rotation about a frontoparallel axis, in which case the POCs lie on a straight line. This straight line is in fact a *trajectory*, that is, it consists of images of a single, fixed point on the object's surface. This trajectory is therefore independent of an object's shape. Using three views of two such fixed points, the object's angular velocity can be computed (Hoffman & Bennett, 1986); the axis of rotation is parallel to the image plane because the trajectories of the fixed points are straight; the axis is perpendicular to these trajectories.

Rieger (1986) also analyzed the fixed points of an object rotating about a frontoparallel axis. He characterized them as the critical points of the height function along the axis of rotation; the surface normal at these points is parallel to the axis of rotation. He suggested using fixed points to compute the angle of gaze in stereo; the advantage of using fixed points is that there is no correspondence problem to be solved (among the fixed points that is) because, generically, the fixed points will have different heights.

It should be emphasized that the above considerations only hold for orthographic projection. Under perspective projection, there might not be any points of correspondence at all between two successive views. For example, if you are walking towards a sphere, successive rims are circles of latitude which do not intersect (Figure 4.6).

If the points of correspondence are not projections of fixed points, their use in computing object motion is more problematic. Recall that two arbitrary views of an object have at least

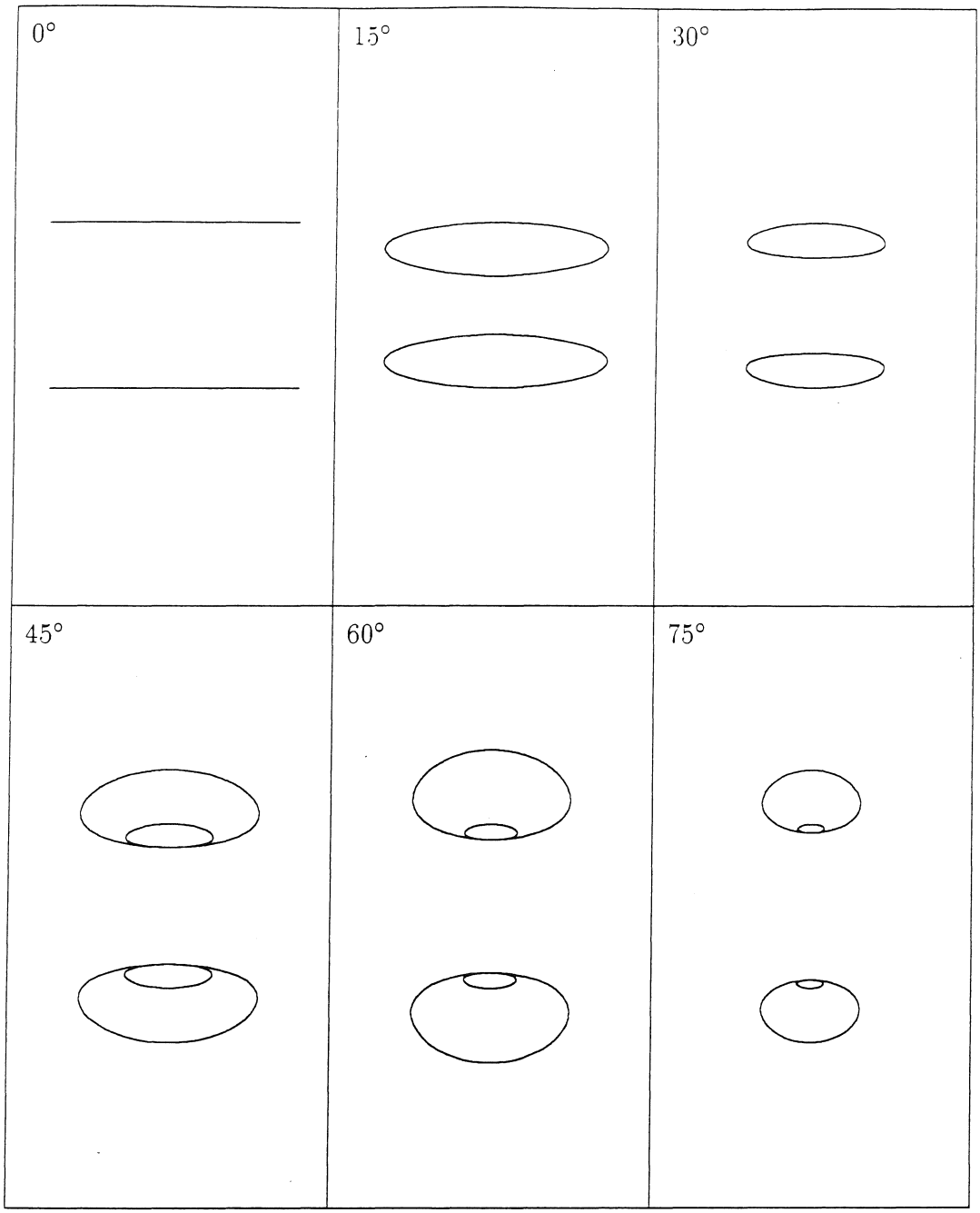


Figure 4.5. Points of correspondence of the ellipsoid $0.625x^2 + 5.000y^2 + 0.167z^2 = 1$, rotated 30° about the x-axis, as it rotates 360° about an axis slanted by various degrees, as indicated in the upper lefthand corner of each panel. The ellipsoid is centered at the origin, and the axis of rotation goes through the ellipsoid's center.

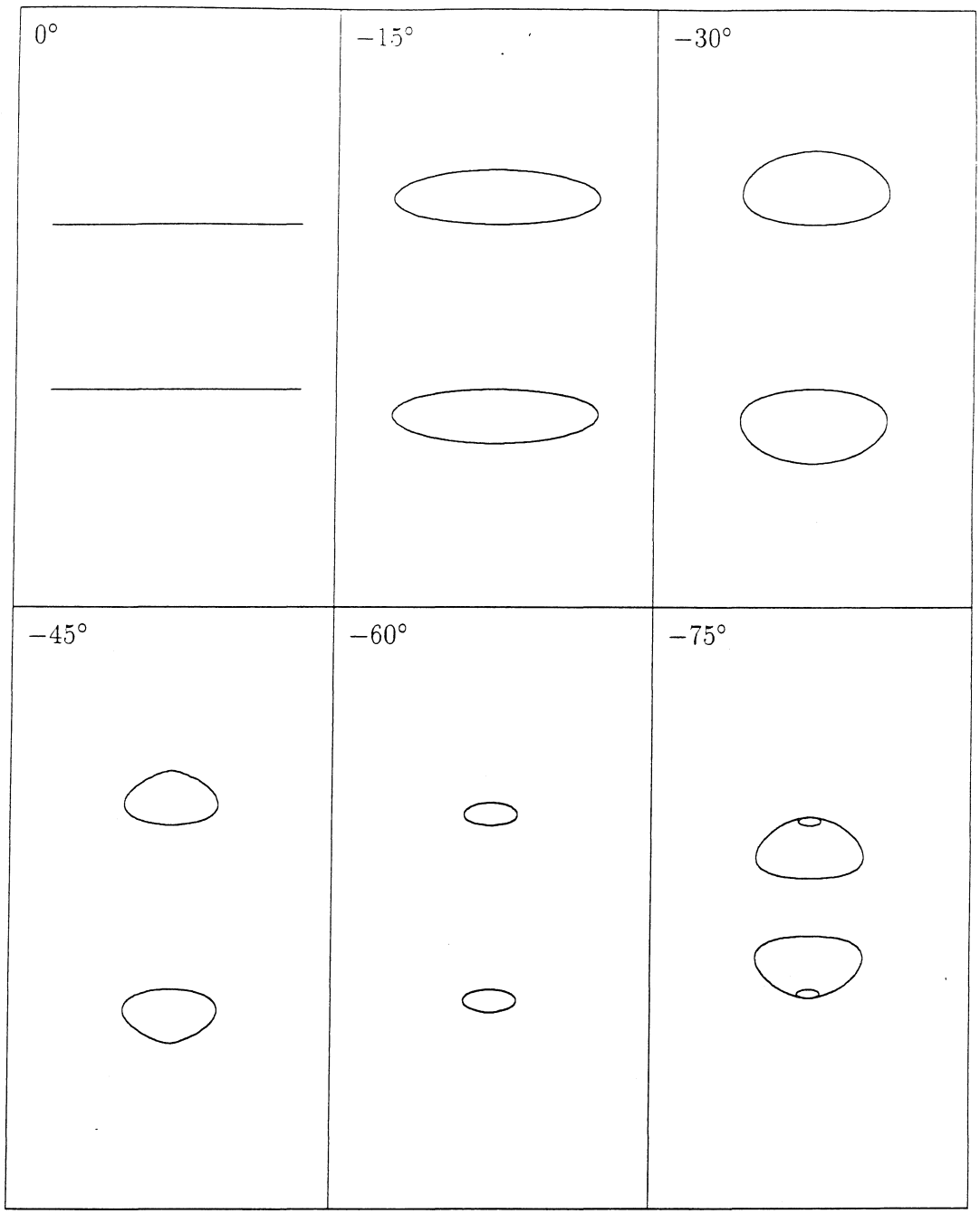


Figure 4.5 cont.

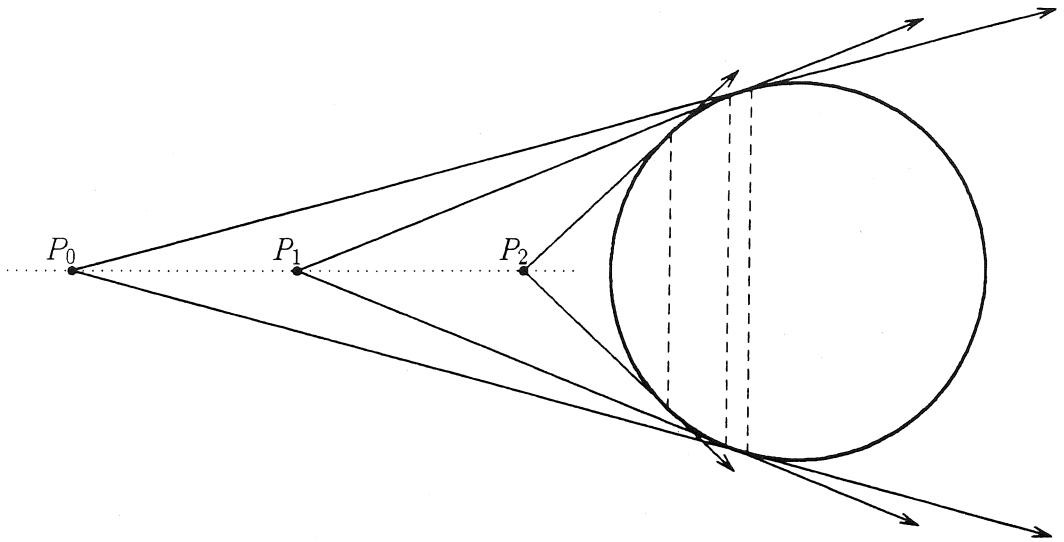


Figure 4.6. Side views of three rims on a sphere (dashed lines) under perspective projection for three vantage points P_0, P_1, and P_2. As the object being viewed is a sphere, these rims can also be interpreted as their Gaussian image on the Gauss sphere. Note that the rims do not intersect. The top angle of the viewing cone associated with P_0 is 28°.

two POCs (under orthographic projection, but not necessarily under perspective projection). Assuming we can locate the POCs of two successive views, can we use them to compute the object's motion? In other words, what can we compute from 2 views of n points, with $n \geq 2$? Bennett, Hoffman, Nicola, & Prakash (1989) showed that two orthographic projections of an arbitrary number of points are compatible with infinitely many interpretations, and that the set of interpretations can be parametrized by the slant of the axis of rotation if there are more than 3 points. Lee & Huang (1989) also showed that two orthographic views have an uncountably infinite number of solutions. This result can be appreciated intuitively if one realizes that any instantaneous rotation about an arbitrary axis can be decomposed into a rotation about the viewing direction and a rotation about a frontoparallel axis. The rotation about the viewing direction can be computed from two views (Ullman, 1979); call the amount of rotation $\sin(\sigma)$. The remaining rotation about a frontoparallel axis, call it $\cos(\sigma)$, is completely undetermined because for each amount of rotation one picks one can adjust the depths of the points such that they yield the observed displacements (this is possible because the points are independent). Hence, the slant of the axis of rotation, which is given by $\operatorname{atan}(\sin(\sigma)/\cos(\sigma))$, is undetermined as well. Successive pairs of snapshots could

be combined to disambiguate the axis of rotation and with it the structure of the object. However, in the present case this is not possible because successive pairs do not have the same POCs and hence cannot constrain each other.

Using perspective instead of orthographic projection, Faugeras & Maybank (1989) showed that 2 views of 5 points have 10 interpretations. Thus, if two perspective views of an object have five or more POCs and if we can locate them, then we could compute the object's motion from just two views.

A related result was derived by Tsai & Huang (1984). Assuming the motion relating the two views is to be expressed in terms of a rotation about an axis through the projection center followed by a translation (which is always possible according to Chasles' theorem), they showed that with 2 views of 8 points rotation and translation can be computed up to a scale factor for the translation; consequently, depth of the points can only be computed up to a scale factor. On the basis of this result, Weng, Huang, & Ahuja (1989) developed an algorithm to compute motion and depth from two views and analyzed in detail its sensitivity to noise. They chose between 10 and 20 points at random from a cube whose edges measured 10 units, viewed them from a distance of 11 units (center of cube to center of projection is 11 units), with the viewing plane at 1 unit in front of the center of projection. One of their main conclusions was that translation has to be quite large in order to compute its direction and the relative depth in the scene with any degree of accuracy. For example, if the translation vector $\mathbf{T}$ is $(1.7, 0, 1.7)$, the relative error in estimating $\mathbf{T}$ and depth is about 40%; only if $|\mathbf{T}| > 3$, does the relative error go down to 10%. In other words, error is considerable even for large displacements.

Summarizing, if we were to use two perspective views we would need quite a few POCs, namely five to use Faugeras & Maybank's (1989) result, and 8 when using Tsai & Huang's (1984) approach (in practice more then 12 were required to get reasonable results), and there would have to be considerable image motion to yield reasonable object motion estimates. Assuming the object persists in its motion, the computations of successive pairs of views can be combined to improve performance. Although not necessary, one would like a computation to degrade gracefully or not at all in going from perspective to orthographic projection as perspective effects become small with increasing viewing distance. Given these considerations and the following psychophysical results, I am skeptical about the possible use of two perspective views to compute object motion.

First, Koenderink & van Doorn (1975, 1986) showed that, theoretically, motion parallax yields the shape of an object modulo a so-called relief transformation, $(\rho_0/\rho)^a$, where ρ is distance to the surface in a particular viewing direction and $\rho_0 > 0$ and a are arbitrary constants; the smaller a, the shallower the relief. (This can be related to the result of Bennett et al., 1989; using the notation explained in Appendix A.6, we see that the shallower the relief, that is, the smaller the nearness gradient $\mathbf{F}$, the larger $\mathbf{A}_t$ has to be in order to produce the same deformation component of motion parallax. Because R_r remains the same, one can interpret this as a decreasing slant of the instantaneous axis of rotation.) Two stereo images can be analyzed in the same manner, giving rise to the same ambiguity. Indeed, Koenderink & van Doorn (1986) note that static stereoscopic views of a relief are virtually indistinguishable from the real object. If the human visual system were using an algorithm based on Faugeras & Maybank's result, then it could be expected to make the distinction.

However, this inability to distinguish between different amounts of relief is consistent with Tsai & Huang's (1984) computational scheme. Summarizing, the human visual system does not seem to use all the information available in two perspective views, perhaps because the resulting computation would be very sensitive to noise.

The second reason for doubting the use of POCs to compute object motion is based on results of Husain, Treue, & Andersen (1989), who showed that the perception of 3D structure becomes less compelling as the lifetime of the points making up the image decreases. If there are only a few points it becomes especially hard to detect a rigid structure let alone determine the axis about which they rotate. For example, in a 2AFC experiment in which subjects had to decide whether or not there was a rigid figure, 77% of responses were correct for images of 126 points on a cylindrical surface, a point lifetime of 100ms, and a display time of 400ms; with only 12 points, performance dropped to about 67%. To the extent that these displays simulate POCs appearing and disappearing during each pair of views, additional doubt is thrown on the suggestion that POCs, of which there are only a few at any one time, are used to compute object motion.

Given the above theoretical and psychophysical results, I do not expect that POCs are in general used by the visual system to determine object motion.

So far I have viewed POCs as isolated from the surface on which they lie. One could argue that if one were to analyze the deformation of the occluding contour in the neighborhood of a POC, one could, at least theoretically, compute an object's motion. This route does not appear to be promising because one needs to know an object's motion in order to extract shape information from its deforming contours (section 4.2).

Having discussed the potential utility of POCs and in particular fixed points, the question remains whether we can find them given a sequence of contours. As the Gaussian image of the pre-image of POCs (i.e., $N(\pi^{-1}(POC))$) lies on the envelope of the Gaussian image of successive rims, it makes sense to investigate the relationship between POCs and the envelope of the occluding contours themselves. Figure 4.7 shows the envelope of a rotating ellipsoid whose axis of rotation is slanted by various degrees (see also Figure 6.9); compare this with Figure 4.5, which shows the POCs for the same ellipsoid rotating about the same axes. The two are obviously very different.

The following two examples will demonstrate that there is no simple one-to-one relationship between POCs and points on the envelope. First consider an ellipsoid rotating about the vertical; the panel labeled 0° in Figure 4.7 shows that the envelope consists of two horizontal lines and a pair of curves in the middle. Because the two horizontal lines are the trajectories of the two fixed points, we can conclude that the envelope of the projection of a rotating solid object does not solely consist of POCs. It might still be that POCs are necessarily part of the envelope. The next example shows this to be false. Consider an ellipsoid rotating about the line of sight; the envelope of its projection consists of two circles whose diameters equal the major and minor axis of the ellipse that is the ellipsoid's projection (section A.4). In other words, the envelope consists of the trajectories of the four points on the ellipse whose distance to the ellipse's center is either maximal or minimal. In fact, however, each point on the contour is the image of a fixed point, which means that fixed points do not necessarily project onto an object's envelope. Summarizing, there is no simple relationship between POCs and the envelope of an object's projection.

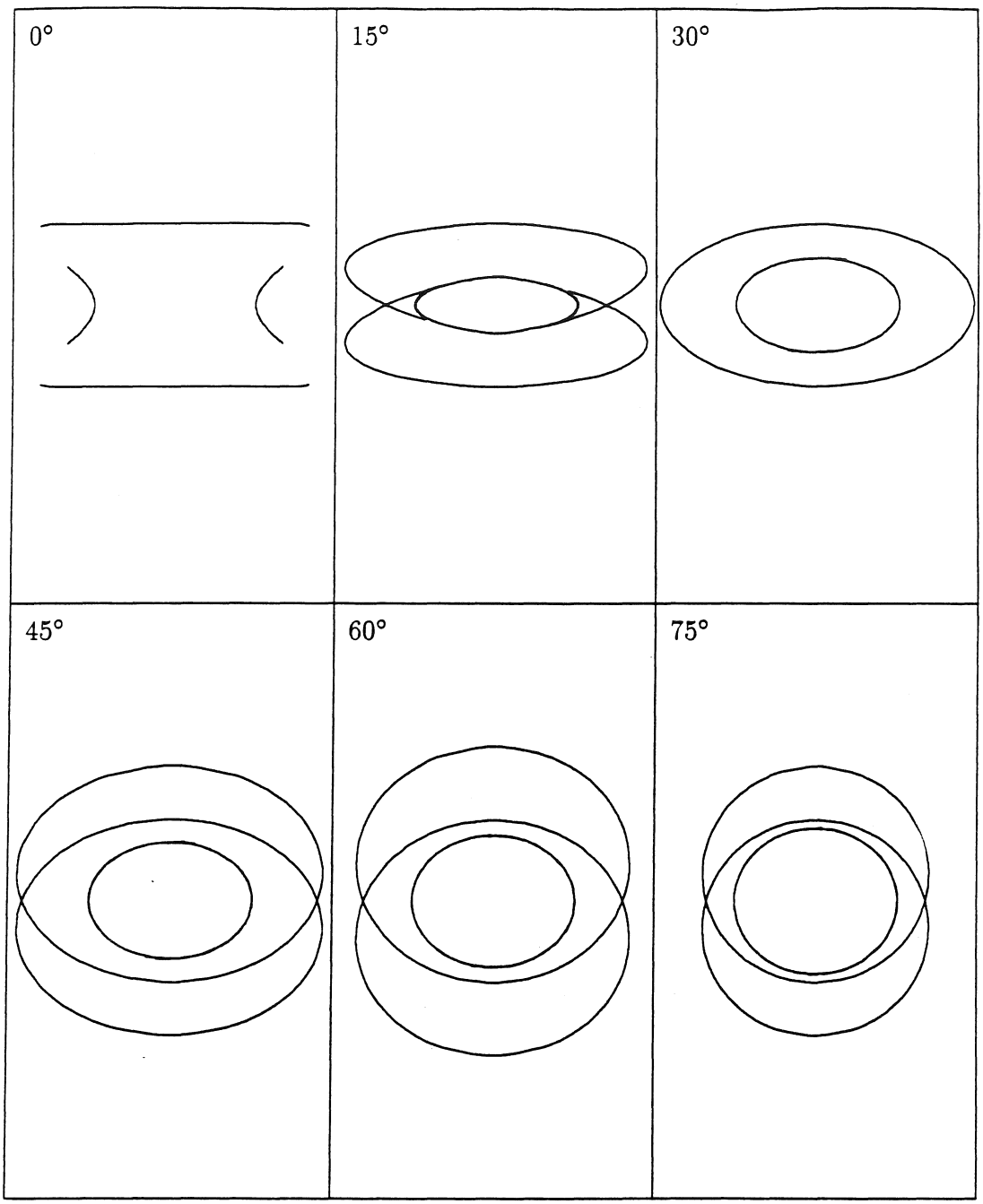


Figure 4.7. Envelopes of the ellipsoid $.625x^2 + 5y^2 + .167z^2 = 1$, rotated 30° about the x-axis, as it rotates 360° about an axis slanted by various degrees, as indicated in the upper left-hand corner of each panel. The ellipsoids are centered at the origin and the axis of rotation goes through the origin.

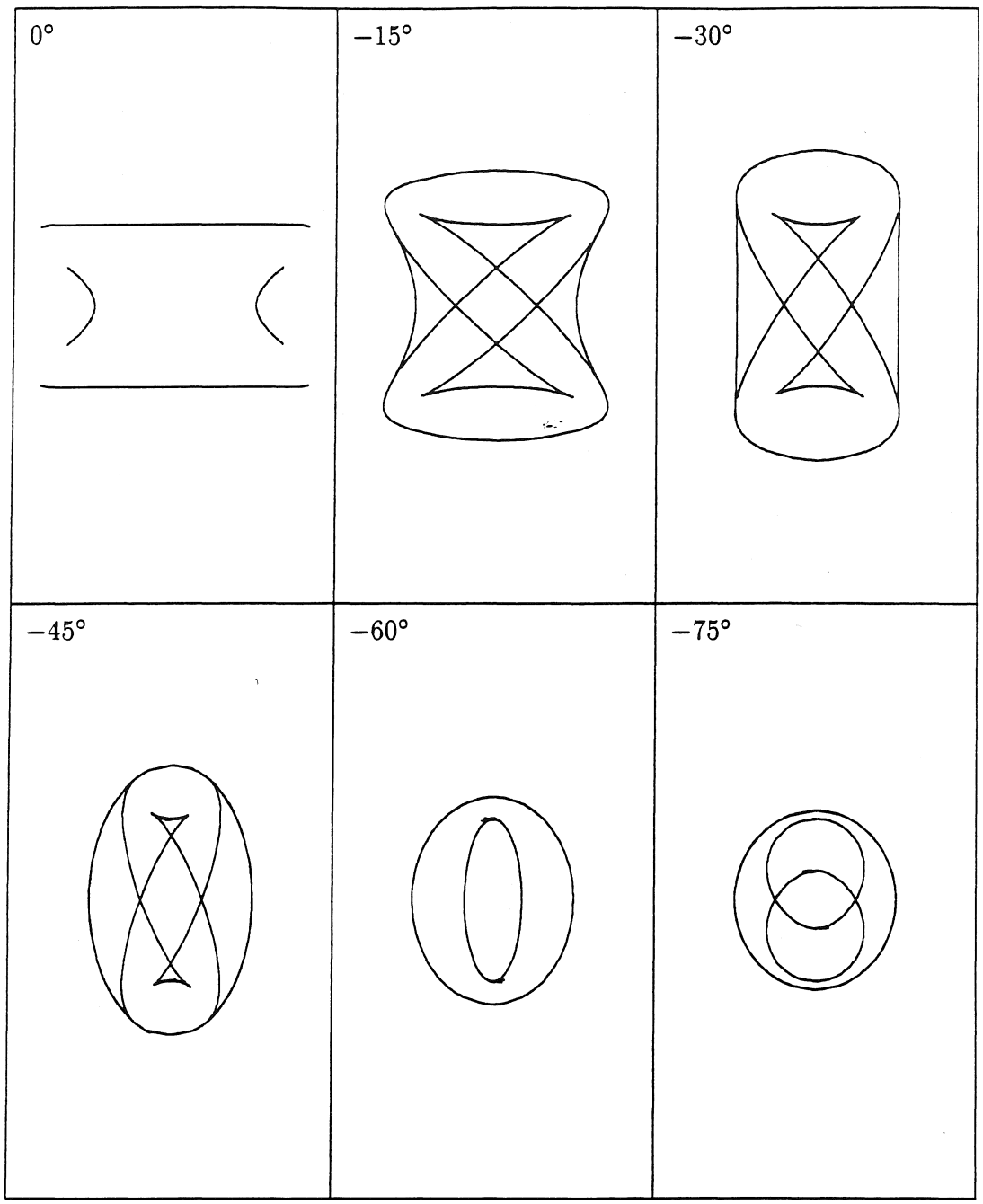


Figure 4.7 cont.

It is still possible, however, to find POCs. Instantaneously, rotation about an arbitrary axis can be decomposed into rotation about the line of sight ($\Omega_\perp$) and rotation about a frontoparallel axis ($\Omega_\parallel$). Let $\mathbf{T}_{ij}$ be a line tangent to contours v_i and v_j, a so-called bitangent. Now, $\Omega_\parallel$ gives rise to at least two bitangents, namely at the POCs; these bitangents are all parallel. As the panel labeled 0° in Figure 4.7 illustrates, there might be additional bitangents; in general, however, these bitangents will not be parallel to each other and hence can be easily discarded. $\Omega_\perp$ disturbs this nice picture by adding a rotation in the image plane which causes the bitangents at the POCs to no longer be parallel. The deviation from parallelism depends on the local curvature of the contour: If the contour at the POCs is curved considerably, the bitangents will remain more or less parallel (Figure 4.8). Given two successive contours v_i and v_j, there exists a rotation $\Omega_\perp$ which will make at least two of the bitangents parallel (namely the ones at POCs); this can be called derotation (Rieger, 1986). Derotating successive contours makes the bitangents at POCs parallel, although they do not necessarily form a straight line; that would only happen if they were images of fixed points. If one has successfully derotated a pair of contours, one has an expectation as to how much the next pair should be derotated (assuming constant angular velocity). Thus we can find the orientation of the axis of rotation in the image plane, but not its slant (how much it is rotated out of the image plane), because we can only connect two successive views. The difference between POCs and the envelope of the projection can now be characterized as follows: The envelope is determined by both $\Omega_\parallel$ and $\Omega_\perp$, whereas POCs factor out $\Omega_\perp$, the rotational component about the line of sight. If $\Omega_\perp = 0$, the POCs will be a subset of the envelope (compare Figures 4.5 and 4.7, panels labeled 0°).

If the slant of the axis of rotation is small, the POCs lie on curves that are virtually straight for considerable stretches (Figure 4.5), and the bitangents might still be used to find them without using derotation.

The bitangency procedure for locating POCs, especially POCs that are images of fixed points, and their subsequent use in computing the object's angular velocity can explain a number of perceptual phenomena. For example, the contour of an ellipsoid rotating about the vertical most often elicits a percept of 3D rigidity if its major axis is not horizontal, that is, not perpendicular to the axis of rotation (Todd, 1985). If the major axis is perpendicular to the vertical and the vertical goes through the ellipsoid's center, then the points of bitangency are stationary. Hence, the object's motion cannot be recovered and the contours cannot be given a 3D interpretation.

The bitangency procedure can also explain some of Todd's (1985) results. Todd found that the silhouette of a blimp rotating about the vertical is perceived as 3D and rigid. The blimp consisted of two prolate ellipsoids, a big horizontal one with a smaller vertical one as a tail attached to it (Figure 2.12). If one separates the two ellipsoids such that the tail no longer intersects the horizontal ellipsoid, and if one looks at the silhouettes of the two ellipsoids rotating in tandem about the vertical, one no longer perceives a 3D solid object from the silhouette of the big, horizontal ellipsoid alone. The points of bitangency of the horizontal ellipsoid are stationary and hence uninformative as we saw in the previous example. The points of bitangency of the vertical ellipsoid both trace a straight horizontal line, thus allowing for the recovery of the axis of rotation and angular velocity. But because the two ellipsoids are not connected, the human visual system apparently does not propagate

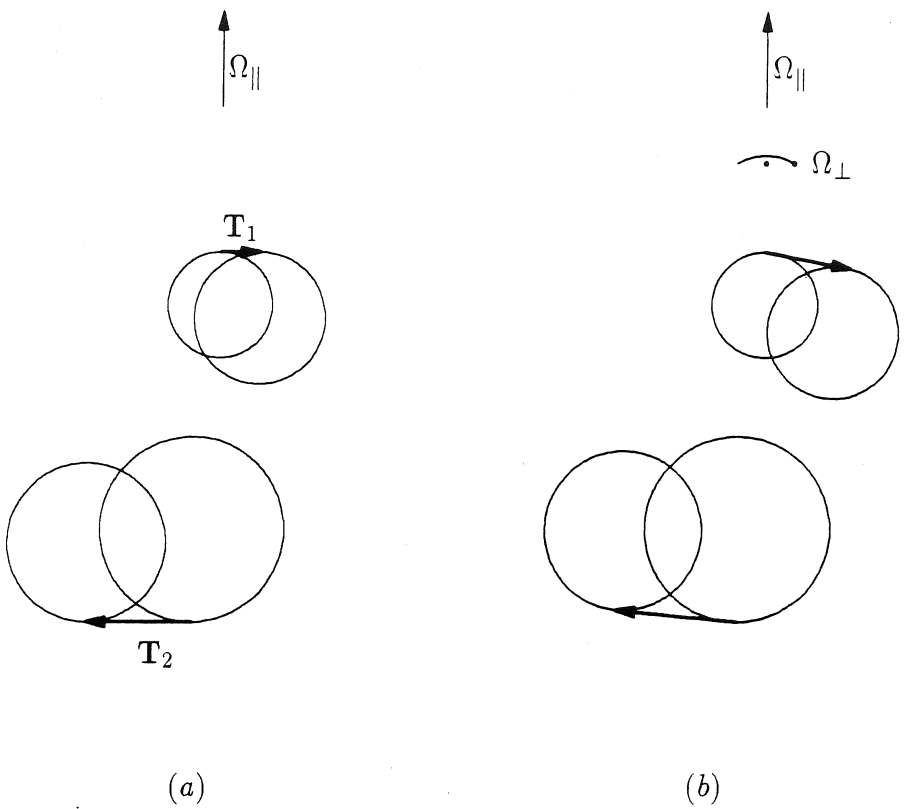


Figure 4.8. The behavior of bitangents under various rotations. The contour at the bitangent is approximated by a circle with radius $1/\kappa_t$, with κ_t the local transverse curvature of the contour. Instantaneously, $\Omega = \Omega_{\|} + \Omega_{\perp}$, that is, rotation about a frontoparallel axis and an axis perpendicular to the image plane. (a) Rotation about $\Omega_{\|}$ only: Bitangents are parallel to each other and perpendicular to $\Omega_{\|}$. Note that $|\mathbf{T}_1| \neq |\mathbf{T}_2|$; this would only be the case if local depth were the same at these points of bitangency. (b) Rotation about $\Omega_{\perp}$ and $\Omega_{\|}$: Bitangents are no longer parallel.

this information to the silhouette of the horizontal ellipsoid, which therefore looks like a 2D deforming blob.

A final example pertains to the so-called barber pole illusion. A barber pole is a cylinder with a spiral painted on it; if the cylinder rotates about its length axis, which has to be frontoparallel, one perceives motion in the direction of the axis even though motion is actually in a direction perpendicular to the cylinder's length axis. In other words, the perception of motion is illusory. However, as soon as one replicates the barber pole's motion on a CRT screen (e.g., Hildreth, 1984), it is no longer correct to refer to the misperception of the direction of motion as illusory, in the sense of perceiving something that has no direct correlate in the image. For example, Mach bands are illusory in this sense because they do not correspond to changes in light intensity in the image. (Ratliff, 1972, relates the story of physicists who, after the discovery of X rays, tried to use ordinary diffraction slits to measure their wavelength. The diffraction patterns they thought to have seen turned out to have been Mach bands!) The barber pole stimulus as generated on a CRT screen is compatible with a rotating 3D spiral as well as with a translating sinusoid. Let $\mathcal{S} \doteq (a\cos\phi, a\sin\phi, b\phi)$ with $-\phi_0 \leq \phi \leq \phi_0$ be a spiral. Rotating $\mathcal{S}$ about the z-axis by $\delta\phi$ yields $\mathcal{S}_R = (a\cos(\phi+\delta\phi), a\sin(\phi+\delta\phi), b\phi)$. Translating $\mathcal{S}$ along the z-axis by $-\delta\phi$ yields $\mathcal{S}_T = \big(a\cos\phi, a\sin\phi, b(\phi-\delta\phi)\big) = (a\cos(\phi'+\delta\phi), \sin(\phi'+\delta\phi), b\phi')$. If we let $-\phi_0 \leq \phi' \leq \phi_0$, then $\mathcal{S}_R = \mathcal{S}_T$. In other words, perceiving the clipped projection of a rotating barber pole as translatory motion of a sinusoid is not illusory. Returning to the bitangency procedure, it is obvious that it will lead to the perception of translation along the axis of the spiral.

One other type of distinguished point on an object's contour should at least be mentioned as a possible source of information about an object's motion. Cusps are easily identified points on the contour and as such can be traced over time. Because their motion depends on local surface shape as well as object motion, they cannot be used to recover motion in an obvious manner.

I conclude that it is hard to determine an object's motion from its smooth contours, that is, contours that do not have any identifiable feature points that could be traced over time. Only when the object rotates about a frontoparallel axis is it possible to determine its axis of rotation and angular velocity in a straightforward manner, namely by using its fixed points. Experiments in sections 6.1 and 6.2 show that some human observers could correctly perceive the angular velocity of ellipsoids rotating about the vertical from their occluding contours. These observers could also discriminate between flat ellipses and solid ellipsoids. For some slants of the axis of rotation, they could no longer discriminate the two. Nor they did perceive the actual axis of rotation; in fact, there was a tendency to see it as frontoparallel.

4.2 Computing shape from occluding contours

In this section, I will focus on the recovery of shape information from the occluding contours of a smooth, solid object rotating about a frontoparallel axis. Obviously, this is a special case. However, I believe that it warrants special attention because it is the only condition in which there are fixed points which can be used to compute angular velocity. As I will argue in this section, angular velocity cannot be computed from points on a contour that are not fixed points.

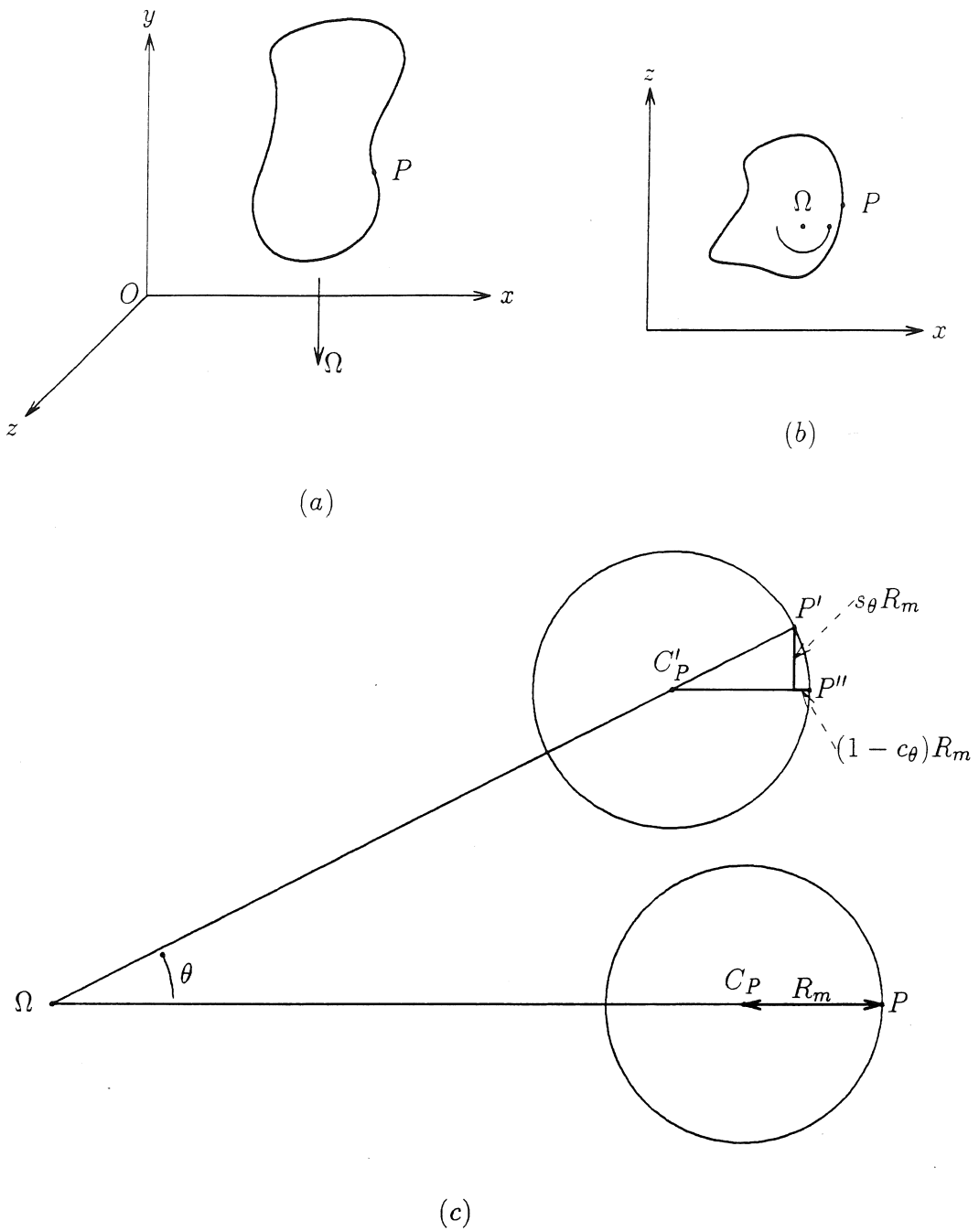


Figure 4.9. (a) A solid object rotating about Ω. (b) A cross section parallel to the $x - z$ plane. As the viewing direction is along the z-axis, P is on the current rim. (c) The radius of curvature of the cross section at P is R_m. After the object rotates θ degrees, P has moved to P' and no longer lies on the rim. The point on the rim is now $P'' = P' + ((1 - c_\theta)R_m, 0, -s_\theta R_m)$.

Consider the situation depicted in Figure 4.9a: A smooth object rotates at constant angular velocity about Ω, an axis parallel to the y-axis: it is viewed under orthographic projection with the $x - y$ plane as the image plane. If we section the object by some plane $y = y_c$, we obtain Figure 4.9b. At the point P on the rim, the cross section can be approximated by a circle with radius $R_m(P) = 1/\kappa_{m(P)}$ and center $C(P)$ as illustrated in Figure 4.9c. The subscript m stands for "motion," as the cross section is taken in the direction of image motion at $\pi(P)$, the projection of P. After a rotation of θ degrees, P will have moved to P'. However, P' no longer lies on the rim. The point on the rim is now P'', which can be obtained by rotating P' about $C(P)$ by $-\theta$ degrees. Thus, the coordinates of P'' are given by:

$$\begin{aligned} x'' &= x\cos\theta - z\sin\theta + R_m(1-\cos\theta), \\ y'' &= y, \\ z'' &= z\cos\theta + x\sin\theta - R_m\sin\theta. \end{aligned} \tag{4.11}$$

It follows immediately that,

$$\begin{aligned} \frac{\partial x''}{\partial\theta} &= -z'', \\ \frac{\partial z''}{\partial\theta} &= x'' - R_m, \\ \frac{\partial^2 x''}{\partial\theta^2} &= -x'' + R_m, \\ \frac{\partial^2 z''}{\partial\theta^2} &= -z''. \end{aligned} \tag{4.12}$$

Thus, the instantaneous velocity field, $\partial x''/\partial\theta$, along the contour is proportional to local depth regardless of surface curvature at the rim. In other words, it does not matter whether we are looking at the occluding contours of a solid object, the projection of a wire frame, or the projection of the edges of a polyhedron such as a cube, relative depth is always proportional to local velocity. Note that $R_m = \partial^2 x''/\partial\theta^2 + x''$; that is, we can compute the radius of curvature R_m from the change in image velocity at P' and the position of P'. (Giblin & Weiss, 1987, first derived these relationships, but in a very different manner.)

The question I want to address in the remainder of this section is whether it is possible to somehow compute angular velocity from points on the contour that are not POCs.

One approach would be to follow a particular location (defined by $y = y_0$) on the contour and measure the local image velocity at time t_1 and t_2. Assume that velocity is measured as a displacement during time dt and that $t_2 - t_1 = ndt$; during the time dt the amount of rotation is $d\theta$, and from t_1 to t_2 it is $nd\theta$. Assuming furthermore that R_m is constant during this time interval, and using a second-order approximation for sine and cosine in Equation 4.11 (as we can only consider small angles $n\,dt$ because of our assumption that R_m remains constant, this second-order approximation is altogether adequate), we obtain:

$$\begin{aligned} x_2 &= x_1\bigl(1 - 1/2(n\,d\theta)^2\bigr) - n\,d\theta z_1 + 1/2(n\,d\theta)^2 R_m, \\ z_2 &= z_1\bigl(1 - 1/2(n\,d\theta)^2\bigr) + n\,d\theta x_1 - n\,d\theta R_m. \end{aligned} \tag{4.13}$$

Substituting $z_1 = -dx_1/d\theta$ and $z_2 = -dx_2/d\theta$, and multiplying both sides of the second equation by $d\theta$, we obtain:

$$\begin{aligned} x_2 &= x_1(1 - 1/2(n\,d\theta)^2) + n\,dx_1 + 1/2(n\,d\theta)^2 R_m \\ dx_2 &= dx_1(1 - 1/2(n\,d\theta)^2) - n(d\theta)^2 x_1 + n(d\theta)^2 R_m, \end{aligned} \tag{4.14}$$

which are two equations in the two unknowns R_m and $d\theta$. Solving for $d\theta$ we get:

$$d\theta = 2\sqrt{\frac{x_2 - x_1 - n(dx_2 + dx_1)/2}{n^3 dx_1}} \tag{4.15}$$

I tried this expression on synthetic input data simulating a sphere rotating about an axis which does not go through its center (otherwise the denominator of Equation 4.15 would vanish); I chose $d\theta = 1°$ and used spheres of different radii. I found that with $n > 7$, Equation 4.15 was accurate to within about 10%. However, the smallest amount of noise in the image data caused the results to become nonsensical; the same happened with slightly inaccurate values for n. Coupled with the fact that R_m is not actually constant, these results indicate that this approach is not viable.

The alternative is to use more than one point. Unfortunately, it is not possible to formulate a constraint among the points, as can be done for points that are rigidly connected (e.g., Ullman, 1979; Hoffman & Bennett, 1985, 1986), or for points that can bend with respect to each other (Koenderink & van Doorn, 1986). Consider, for example, three neighboring points on a contour: Q_1, Q_2, and Q_3 defined by $Q_i = y_i$, where the y_i are constants. Over time the contour changes and with it the x-coordinates of the $Q_i(t)$. Let $P_i(t) = \pi^{-1}(Q_i(t))$, that is, P_i is the pre-image of Q_i. First, note that, generically, there is no correspondence between views: $P_i(t_1) \neq P_i(t_2) \neq P_i(t_3)$ (except when P_i is a fixed point; but then the neighbors of $Q_i = \pi(P_i)$ cannot be images of fixed points too because images of fixed points are isolated points on a contour). It might still be that the 3D distance between points does not change over time, that is, $d\big(P_i(t), P_j(t)\big) = d\big(P_i(t+1), P_j(t+1)\big)$. I can think of only one case in which this is true, namely if the object we are looking at is a sphere. Because the rim of a sphere always is a circle of the same radius, the 3D distances among the P_i's remain the same. In general, however, the rim changes over time and with it the 3D distances. And because the distance between the P_i's as measured along the rim changes over time, the deformation of a rim over time cannot be modeled as bending.

Points on the rim are not completely unrelated however. The direction of the rim and the viewing direction are conjugate directions (section 2.1). But this does not provide us with a constraint across the different views because different views are from different parts of the object each with its own conjugate direction.

I conclude that it is not possible to compute angular velocity from a few arbitrary points on an object's contour. It is of course possible to apply an algorithm embodying the 3D rigidity constraint to occluding contours. If the computed angular velocity is different for different parts of the contour, one can conclude that one is looking at something nonrigid or the contour of a rigid object.

The following is an example of such an algorithm. It combines two consecutive velocity measurements and assumes that the object is rotating at constant angular velocity. Each

velocity measurement yields partial information (i.e., relative depth confounded with angular velocity, see Equation 4.12); by combining two such measurements from sufficiently different views, complete information can be obtained.

Consider two neighboring points P_1 and P_2 rotating about the y-axis. Let $P_2 = P_1 + (dx, dy, dz)$ at time t_1, and $P_2' = P_1' + (dx', dy', dz')$ at t_2. The velocity difference between P_2 and P_1 is measured during time t_0 to t_1, during which time the object rotated $d\theta$ degrees: $dv = dz\, d\theta$. Similarly, the velocity difference between P_2' and P_1' is measured during time t_2 to t_3: $dv' = dz'\, d\theta$. From time t_1 to t_2 the object rotated θ degrees, so that

$$\begin{aligned} dx' &= \cos\theta\, dx - \sin\theta\, dz, \\ dy' &= dy, \\ dz' &= \cos\theta\, dz + \sin\theta\, dx. \end{aligned} \tag{4.16}$$

Multiply both sides of the above equations by $d\theta$, and substitute $d\theta\, dz = dv$ and $d\theta\, dz' = dv'$ in order to obtain equations in terms of quantities measurable in the image. This yields two equations in the two unknowns $d\theta$ and $\cos\theta$:

$$\begin{aligned} dx' &= \cos\theta\, dx - \frac{dv}{d\theta}(1 - \cos^2\theta)^{1/2}, \\ dv' &= \cos\theta\, dv + d\theta(1 - \cos^2\theta)^{1/2} dx. \end{aligned} \tag{4.17}$$

Solving for $\cos\theta$ gives:

$$\begin{aligned} \cos\theta &= \frac{dv\, dx + dv'\, dx'}{dv\, dx' + dv'\, dx} \\ &= \frac{dv \cot\alpha_1 + dv' \cot\alpha_2}{dv \cot\alpha_2 + dv' \cot\alpha_1}, \end{aligned} \tag{4.18}$$

where $\cot\alpha_1 = dx/dy$ and $\cot\alpha_2 = dx'/dy'$. Knowing θ, we can compute dz and dz' from Equation 4.16. Note that Equation 4.18 contains a consistency check on the image data because $|cos\theta| \leq 1$. I will use this algorithm in section 6.2 to compute angular velocity using occluding contours of rotating ellipsoids.

4.3 Models of the perception of shape from contours

Assuming for the moment that human observers actually perceive a rigid, solid object when looking at the occluding contours of a rotating smooth object, we can ask what kind of model they entertain. In the next three sections, I outline three different models; they are illustrated in Figure 4.10. In Chapter 6, I will elaborate upon these models in the context of a particular class of objects, namely ellipsoids, and a particular psychophysical task, the discrimination of solid ellipsoids from planar ellipses.

4.3.1 Rim prediction model

The basic idea is that an important aspect of an internal model of an external object is its ability to predict future visual input, in this case the future appearance of an object. In other words, the current visual input sets up expectations regarding the outcome of subsequent

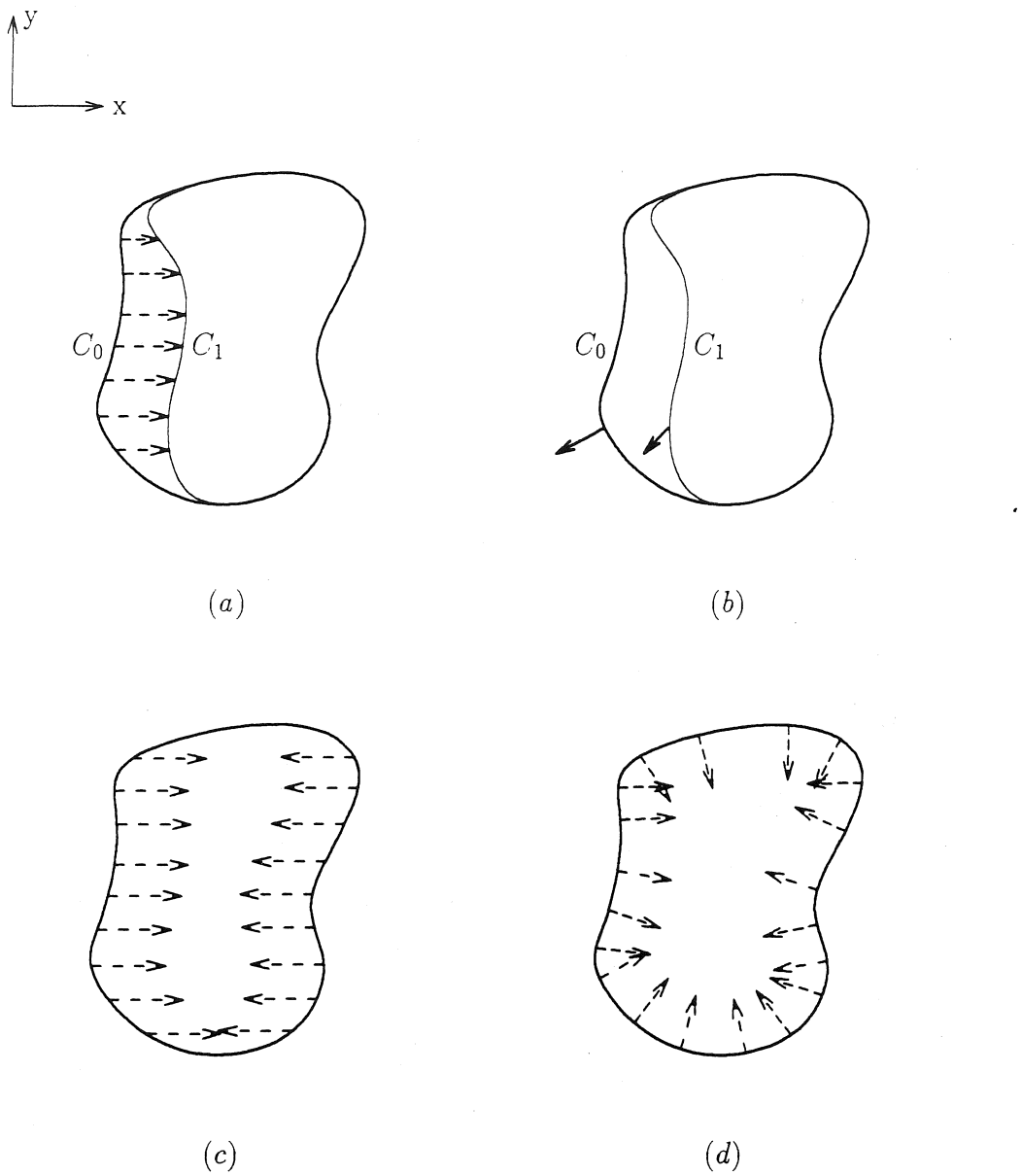


Figure 4.10. Models for the perception of solid shape from dynamic contours; object rotates about an axis parallel to the y-axis. (a) C_0 is the contour at time t_0; along $\pi^{-1}(C_0)$ relative depth is $z = -dx/d\theta$. This is used to predict depth along C_1, the projection of $\pi^{-1}(C_0)$ at time t_1 had it been a wire figure. (b) In addition to, or instead of, depth along C_1 we can predict the surface normal or tangent plane along C_1. Alternatively, we can use the contour's deformation to compute a measure of surface curvature such as radial curvature or Gaussian curvature, and propagate it towards the inside of the figure. Propagation can be in the direction of motion (c), or in the direction of the local normal to the contour (d).

interactions between observer and environment. This is a way to afford an observer the feeling of continuity over time, to link the present with the future.

Some interesting examples are given by Koenderink & van Doorn (1982b). Just as the direction of the rim is conjugated to the viewing direction, so is the direction of an isophote conjugated to the direction of the light source (isophotes are curves on an object that project onto curves in the image along which light intensity is constant). Under perspective projection and for surfaces whose specular reflection is small relative to its diffuse reflection, isophotes are potential rims and vice versa. (This relationship between rims and isophotes follows immediately if one considers the Gaussian image of an object. On the Gauss sphere, the viewing direction is locally perpendicular to a rim's Gaussian image and to that of an isophote. Orthogonal directions on the Gauss sphere correspond to conjugate directions on the object's surface.) If an observer moves relative to some textured object, the contour lines of the divergence of the optic flow field are future rims (for objects that are sufficiently curved that is; Koenderink & van Doorn, 1975). For example, if one looks at a textured cylinder rotating about its frontoparallel length axis, the contour lines of divergence are parallel to the cylinder's length axis. Because divergence equals
$-\mathbf{F} \cdot \mathbf{A}_t + 2A_r$ (section A.6), and because the variation of $\mathbf{A}_t$ and A_r over the cylinder's image is small compared to that of $\mathbf{F}$, the contour lines of divergence are the lines of equal surface slant $\mathbf{F}$, which are vertical lines. This approximation is not true for a frontoparallel plane rotating about the vertical; now the variation in $\mathbf{F}$ is on the order of the variation in $\mathbf{A}_t$ and A_r.

For the present case of an observer looking at deforming contours in the image, the observer computes relative depth along the contour and uses it to predict subsequent rims or contours. If predicted and actual visual input agree, the observer perceives a rigid object in this case in the form of a 3D wire. If the two disagree, as they would for the occluding contours of a smooth object, the observer either perceives a nonrigid wire or a rigid, solid object whose internal structure (i.e., relative depth) derives from past views.

4.3.2 Tangent plane prediction model

At a point P on a contour, be it the projection of a wire or of a solid object, the tangent plane is known: It is spanned by the local tangent to the contour at P and the local viewing direction. Together with knowledge of relative depth along the contour, we can predict future locations and orientations of this tangent plane. Suppose we are looking at a rotating wire; the predicted and actual locations of future tangent planes coincide, whereas the predicted and actual orientations of the tangent plane will differ. This signals infinite curvature (in the direction of motion) as is in fact true for the wire. If we are looking at a smooth solid object, both the location and orientation of predicted and actual tangent planes will differ, which corresponds to a certain amount of surface curvature.

4.3.3 Curvature model

The observer directly computes R_m, the radius of curvature in the direction of motion, along the contour. Knowing R_m it is possible to compute radial curvature, Gaussian curvature, and mean curvature (Giblin & Weiss, 1987). Surface curvature information is then propagated inwards (i.e., towards the figure side of the contour), giving rise to a percept of solid

shape. Beusmans, Hoffman, & Bennett (1987) and Richards, Koenderink, & Hoffman (1987) studied the propagation of qualitative shape information, in the form of the sign of Gaussian curvature, in static contours. In particular, they looked at the number of ways in which this shape information can be propagated and still yield a globally consistent curvature interpretation, and at possible constraints to limit this potentially very large number of consistent interpretations.

One consideration which makes the curvature model implausible, at least for human visual perception, is that directly computing R_m from the deformation of the contour requires accurate knowledge of x, that is, the distance between a point and the axis of rotation, because $R_m = x + \partial^2 x / \partial \theta^2$. The second-order derivative also promises sensitivity to noise. Perhaps surface curvature is computed indirectly from changes in the tangent plane as described in section 4.3.2.

4.4 Summary

In this chapter, I studied the inference of solid shape and object motion from the occluding contours of a rotating object. I divided this process into three successive stages. In the first stage, the motion of the object is determined, that is, its axis of rotation and its angular velocity. I argued that only if the object is rotating about a frontoparallel axis is it straightforward to determine its motion. In that case one can use the images of fixed points to compute angular velocity and the orientation of the axis of rotation in the image. In the second stage, relative depth and surface curvature are computed along the rim. In the third and final stage, these local measurements are integrated into a global percept of solid shape. I formulated three models for this integration process. The rim and tangent plane prediction models assume knowledge of relative depth along the rim and of the position of points on the contour relative to the axis of rotation. In the curvature model, surface curvature is computed directly from the deforming contour, which also requires knowing the position of points on the contour relative to the axis of rotation. In Chapter 6, I will study how well the different models can predict the performance of human subjects in discriminating solid ellipsoids from planar ellipses.

A serious drawback of all three models is the need to know a point's position in the image relative to the axis of rotation, although it seems that the curvature model would require more accurate knowledge (see section 6.1 for computer simulations supporting this claim). Perhaps the human visual system simply assumes correspondence and uses any of a number of "structure-from-motion" algorithms (e.g., Equation 4.18) to compute 3D structure. This only requires knowing the positions of small numbers of points relative to each other. For occluding contours the assumption is of course wrong; failure can be detected through consistency checks (see section 6.2).

Chapter 5
Visual perception of surface orientation

The initial motivation for the studies in this chapter was the observation that the slant of ellipses and ellipsoids rotating about the vertical is underestimated (Chapter 6). As I used orthographic projection to compute the projection of these objects, this underestimation could be attributed to the absence of perspective transformations. The use of perspective projection, however, did not result in a veridical perception of slant for the ellipses and ellipsoids. This might be because there are only two fixed points on an ellipsoid rotating about a frontoparallel axis, points that are separated by a large distance (approximately 6 degrees of visual angle), or because the motion field is hard to recover given that the contours are smooth, i.e., without identifiable feature points. The latter interpretation is supported by the observation that rotating ellipses rendered by clearly separated dots appear to be slanted a little more than when rendered by smooth curves, and the perception of slant is more compelling. Thus, the general question remains as to the importance of perspective in the perception of slant. The results of experiment 5.1 in this chapter suggest that the underestimation of slant for rotating planar surfaces cannot be attributed to the absence of perspective. Furthermore, the outcome of experiment 5.1 hints at certain mechanisms underlying the perception of slant.

Part of the current work on structure from motion (e.g., Bennett, Hoffman, Nicola, & Prakash, 1989; Hoffman, 1982; Hoffman & Bennett, 1985, 1986; Ullman, 1979, 1984) is based on the assumption that the human visual system analyzes an image, at least locally, as if it were obtained through orthographic projection.

Ullman (1979) gave the two main justifications for this assumption. The first asserts that, locally, perspective effects are so small as to be negligible. This may or may not be true: It depends on what one actually measures. Consider, for example, a line segment that rotates in the image. If one measures the displacement of its endpoints, the smaller the segment's length the smaller these displacements will be. But if one were to measure the change in orientation, the length of the segment has no such effect. In the context of a structure-from-motion algorithm in which local image displacements are used, perspective effects are indeed small, but the human visual system might use other algorithms to compute shape. One proposal is based on changes in orientation of line segments, sometimes referred to as orientational disparity (see Koenderink, 1986, and Koenderink & van Doorn, 1976a, for a discussion of the use of orientational disparity to compute the deformation of the optic flow field, and hence surface slant; see Wolff & Boult, 1989, for a computational study of the use of orientational disparity in stereo; for psychophysical evidence concerning the use of orientational disparity in the stereoscopic perception of slant see Ninio, 1985, and Rogers & Koenderink, 1986, for slant perception through optic flow; for a discussion of the results of the search for neurons selective for orientational disparity see Nelson, 1986). Prazdny (1983) showed that the angle between the epipolar lines for the left and right eyes is about 2.5 degrees of visual angle for a fixation distance of 6 interocular distances. In other words,

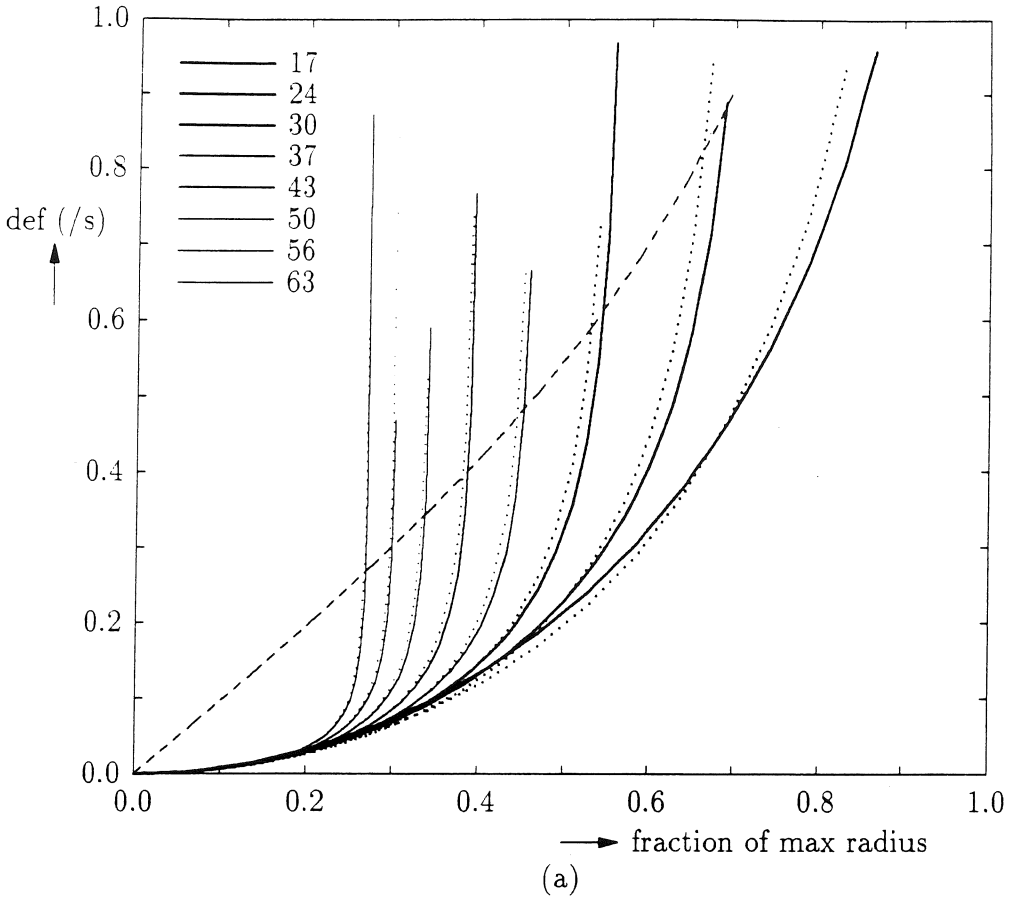
perspective can cause considerable differences in orientation: this orientational disparity is independent of the line segments' length. And Cutting & Millard (1984) point out that many of the objects in whose shape we are interested are in fact quite close to us, making the use of perspective projection appropriate.

The second and main result which Ullman (1979) employed to argue against the use of perspective in computing shape was the fact that human subjects failed to perceive 3D structure in displays simulating two cylinders inside each other translating towards and away from the subject. If subjects had used a perspective analysis they would have been able to compute 3D structure, the argument goes. However, this is a negative result, possibly due to the experimental design. To investigate this possibility, I decided to look at the deformation of the velocity field induced by these stimuli. Deformation is the one first-order differential invariant of the velocity field whose magnitude is proportional to the magnitude of slant (Koenderink, 1986; Koenderink & van Doorn, 1975, 1976b, 1986).

I computed the deformation in the optic flow field between each transition using Ullman's description of the stimuli (Ullman, 1979, p.159; see section A.10 for details; computer simulations of these moving cylinders, using the method outlined in section A.6 to compute differential invariants, confirmed these results). The results are shown in Figure 5.1 for the larger of the two cylinders. Clearly, the deformation is quite small if the distance between the cylinder and observer exceeds 30cm, except of course close to the occluding contour of the cylinder where the deformation becomes infinitely large as it is proportional to the tangent of the angle between viewing direction and surface normal. As an aside, note that there is little difference between the deformation at a point P in the image induced by the cylindrical surface and by the local tangent plane approximation to the cylindrical surface in the direction of P (Figuree 5.1 lines vs. dotted lines, respectively). For comparison, the deformation for a plane slanted by 22° and rotating about the vertical at $17.17°/s$ is $.121/s$ (this is at the moment the projection of the plane's surface normal is parallel to the vertical; as the plane continues to rotate, deformation increases as we obtain more and more a side view of the plane).

I also computed the deformation for the same cylinder rotating at $50°/s$ about its length axis, at a distance of 30cm from the observer (dashed line in Figure 5.1; see section A.11 for details). Ullman (1979; p.134) used up to $100°/s$ which would double deformation as it is proportional to angular velocity. The latter relationship implies that, under rotation, the deformation associated with the smaller cylinder is half that of the bigger cylinder because their radii differ by a factor of two; the difference under translation, on the other hand, is much less. Also note that we would get the same amount of deformation if the cylinder had been translating in the horizontal direction, parallel to the image plane at $30(\text{cm}) \times (50°/s) \times \pi/180 = 26.2\text{cm}/s$ (see section A.8). Thus, the deformation induced by a vertical cylinder rotating about its length axis (under orthographic or perspective projection) or translating in a horizontal direction (under perspective projection) is an order of magnitude larger than that induced by a cylinder translating in depth (under perspective projection).

If the perception of surface slant is determined by the amount of deformation, for which I obtained evidence in experiment 5.1, then the results of Ullman's experiment are not surprising. Subjects failed to perceive 3D structure because deformation was small, perhaps even below threshold, most of the time and for most of the cylinder's image. Only close to



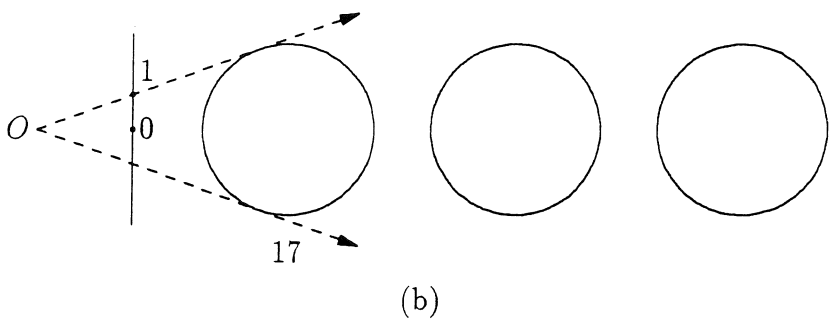


Figure 5.1. (a) Deformation (/s) for a cylinder of radius 2.5cm moving towards the observer at 65cm/s for a number of initial distances in cm (as indicated in the upper left-hand corner); we only consider a horizontal cross section; the abscissa indicates fraction of the distance from the origin to the occluding contour of the closest, and therefore largest, cylinder. Dotted lines refer to the deformation of a plane approximating the local tangent plane to the cylinder. The dashed line refers to a cylinder at a distance of 30cm and rotating at $50°/s$ about its length axis. (b) Top view of observer and cylinder in various positions. 0 and 1 are extrema of the fraction of max radius in image.

the cylinder's occluding contours is the deformation large, indicating 90° slant. The pattern of deformation of this cylinder is indistinguishable from that of a flat box, whose slant is zero everywhere except along its sides. This interpretation is supported by casual observations of a receding flat surface having one horizontal ridge in the form of a cosine between $[-90°, 90°]$. If the height of the ridge is less than half its width, no contours are apparent in the image and the surface looks flat. If contours become apparent due to local self-occlusion, the ridge becomes apparent; however, it looks like a step function instead of a cosinusoid. This was reported by several naive observers. Another possible reason the translating cylinders failed to elicit a 3D percept is the fact that there were two cylinders inside each other; it might be that the two cylinders are easier to separate under rotation than under translation because the image velocities and deformations differ by a factor of two under rotation (because the radii of the two cylinders differ by a factor of two) and only little under translation.

To summarize, if one wants to determine whether or not the human visual system applies a perspective analysis to its visual images one has to use perspective images that are comparable to the orthographic ones in some particular sense. If, for example, one believes deformation to be important in slant perception one has to compare rotation about a frontoparallel axis under orthographic projection with translation by an appropriate amount in a frontoparallel direction under perspective projection rather than translation in depth. Translation in depth induces very little deformation making it impossible to equate it with rotation about a frontoparallel axis. To avoid the problem of some motions being more efficacious in segregating overlying surfaces, one should use opaque objects rather than transparent ones. In the discussion of experiment 5.1, I compare the perception of slant for planar surface patches rotating about a frontoparallel axis with patches translating in a frontoparallel direction.

Before detailing experiment 5.1, I will motivate the use of deformation some more, and I will discuss the difference between the velocity fields of rotating planes under orthographic and perspective projection.

Recent results of Kaiser, Perrone, Andersen, Lappin, & Proffitt (1990) on the perception of surface slant can be interpreted using the amount of deformation in the optic flow field. Kaiser et al. found that translation towards a surface leads to a less accurate perception of its slant than does translation in an orthogonal direction (using perspective projection and with a texture gradient due to the slant). This is not too surprising because translation towards a surface causes very little deformation in the optic flow field (sections A.12 and A.14). Consider, for example, a plane slanted 60° and at a distance of 10 units from the observer, so that it subtends 6° by 6°. If the observer moves one unit away from the plane, divergence ranges from -.16 to -.20, curl is zero, and the maximal deformation is .013. The same displacement in a frontoparallel direction (in this case, perpendicular to the surface's slant), leads to negligible divergence, but to a considerable curl and deformation, both .17. It could very well be that slant judgments for the two directions of motion would become similar if the amount of translation were controlled such that both conditions had the same deformation. So it might not be an effect of the direction of motion per se, as the authors conclude, but a matter of the amount of deformation produced by the different motions.

To get a more quantitative handle on the issue of perspective versus orthographic projection, I derived equations for the velocity field of a slanted plane rotating about a vertical axis

under both kinds of projection (section A.9). Table A.2 shows that the difference between perspective and orthographic projection is on the order of minutes/s for a plane rotating at $15°/s$, and that the difference is about the same in the horizontal (u) and vertical (v) directions, that is, perpendicular and parallel to the axis of rotation. Under orthographic projection, $v = 0$, so one would expect the difference between orthographic and perspective projection to be the most noticeable in the vertical direction. This might explain some psychophysical results reported in Braunstein (1976, pp. 138–139). Subjects had to indicate the direction of rotation of a rectangle rotating about the vertical. Under orthographic projection, this direction is ambiguous. By adding perspective deformations either in the horizontal direction or in the vertical direction, the relative importance of the two can be evaluated. Vertical perspective led to the same performance as complete (vertical and horizontal) perspective. Horizontal perspective by itself did improve accuracy somewhat but performance remained well below the level obtained by vertical or complete perspective. This suggests that the changes in the vertical direction are more noticeable than those in the horizontal direction.

Another way of analyzing the difference between orthographic and perspective projection is by looking at the following differential invariants: divergence, curl, and deformation. Table 5.1 gives some values of these invariants for a plane rotating about a vertical axis. Note that the values in the table only refer to the point on the plane that is straight ahead from the observer. However, curl and deformation hardly vary at all across the plane; divergence varies somewhat under perspective projection because it is the only invariant that depends on A_r, the relative motion between observer and object in the local viewing direction (looming).

Consider first the case in which the axis of rotation is contained within the projection plane (see Figure A.3), that is, $P_d = R_d = 10$ in Table 5.1. The only effect in going from perspective to orthographic projection is that the variation in divergence disappears. For example, with $\Omega_h = \{0°, 30°\}$ and $\Omega_v = 30°$, divergence ranges from -.013 to -.027 in going from left to right across the plane (there is no variation in the vertical direction). This suggests that the human visual system uses the variation in divergence to determine the direction of rotation (Braunstein, 1976).

Moving the axis of rotation out of the projection plane does not affect the invariants under orthographic projection. In fact the position of the projection plane and the axis of rotation are immaterial under orthographic projection. It is straightforward to show that the distance between the orthographic projection of two points rotating about a frontoparallel axis is not affected by the location of that axis; consequently, the differential invariants remain the same. Let $P = (x_p, y_p, z_p)$ and $Q = (x_q, y_q, z_q)$ be two points rotating about an axis $\Omega = (0, 0, 1)$; let the viewing direction be $(0, 1, 0)$, in other words, the xz-plane is the image plane. First consider the case where Ω goes through the origin. Then after a rotation of θ degrees, $P' = (x_p c_\theta + y_p s_\theta, y_p c_\theta - x_p s_\theta, z_p)$ and $Q' = (x_q c_\theta + y_q s_\theta, y_q c_\theta - x_q s_\theta, z_q)$; thus, the projection of $P' - Q'$ equals $d'_{pq} = ((x_p - x_q)c_\theta + (y_p - x_q)s_\theta, z_p - z_q)$. Now let Ω be located at $(0, R_y, 0)$, so that after a rotation of θ degrees, $P'' = (x_p c_\theta + (y_p - R_y)s_\theta, (y_p - R_y)c_\theta - x_p s_\theta + R_y, z_p)$ and $Q'' = (x_q c_\theta + (y_q - R_y)s_\theta, (y_q - R_y)c_\theta - x_q s_\theta + R_y, z_q)$, so that $d''_{pq} = ((x_p - x_q)c_\theta + (y_p - R_y - x_q + R_y)s_\theta, z_p - z_q) = d'_{pq}$.

Table 5.1 Differential invariants of a plane rotating about the vertical under perspective (p) and orthographic projection (o). Without slant and a projection distance of 10 units, the plane subtends 10° by 10°. Divergence, curl, and deformation are given for various degrees of rotation about the horizontal (Ω_h), vertical (Ω_v), for different projection distances (P_d), and positions of the axis of rotation (R_d). The amount of rotation between views is 2°. Values are for the center of plane, that is, in the direction straight ahead from the observer.

				div		curl		def	
R_d	P_d	Ω_h	Ω_v	p	o	p	o	p	o
10	10	0°	0°	0.0	0.0	0.0	0.0	0.0	0.0
		-30°	0°	0.0	0.0	.023	.023	.018	.018
		-60°	0°	0.0	0.0	.062	.062	.059	.059
		0°	30°	-.020	-.020	0.0	0.0	.020	.020
		0°	60°	-.060	-.060	0.0	0.0	.061	.061
		-30°	30°	-.020	-.020	.023	.023	.030	.030
		-30°	60°	-.061	-.061	.041	.041	.073	.073
		-60°	30°	-.020	-.020	.070	.070	.072	.072
		-60°	60°	-.060	-.060	.122	.122	.135	.135
10	7	0°	0°	0.0	0.0	0.0	0.0	0.0	0.0
		-30°	0°	0.0	0.0	.029	.020	.028	.020
		-60°	0°	0.0	0.0	.087	.060	.086	.060
		0°	30°	-.028	-.021	0.0	0.0	.031	.021
		0°	60°	-.088	-.061	0.0	0.0	.088	.061
		-30°	30°	-.031	-.021	.034	.023	.045	.032
		-30°	60°	-.089	-.062	.058	.040	.105	.074
		-60°	30°	-.031	-.021	.101	.070	.105	.073
		-60°	60°	-.090	-.062	.174	.121	.196	.136
7	10	0°	0°	0.0	0.0	0.0	0.0	0.0	0.0
		-30°	0°	0.0	0.0	.015	.020	.014	.020
		-60°	0°	0.0	0.0	.043	.061	.042	.061
		0°	30°	-.014	-.021	0.0	0.0	.015	.021
		0°	60°	-.042	-.061	0.0	0.0	.043	.061
		-30°	30°	-.013	-.021	.017	.023	.022	.031
		-30°	60°	-.043	-.061	.029	.040	.051	.073
		-60°	30°	-.014	-.021	.050	.070	.051	.073
		-60°	60°	-.043	-.061	.085	.121	.095	.135

As expected, the values of the invariants under perspective projection decrease as the axis of rotation gets closer to the observer. If the axis of rotation is at the observer, the motion corresponds to a pure rotation of the eye and all three invariants vanish.

Psychophysical experiments have shown that slant is underestimated in a number of circumstances, including static displays having texture gradients and perspective (Rosinski & Levine, 1976; Perrone, 1982), and dynamic displays simulating surface patches translating along the horizontal (Braunstein, 1968; Braunstein & Tittle, 1988). On the other hand, Braunstein, Tittle, & Liter (1990) reported accurate judgments for planes slanted up to 40° and rotating about the vertical under orthographic projection.

The following experiment extends the range of slants to 70° for rotating planar patches. I compare the resulting slant judgments with those for translating planes under perspective projection.

Experiment 5.1. Judging the slant of rotating planar surface patches.

The purpose of this experiment is to determine how accurately the orientation of a planar patch is perceived under orthographic and perspective projection for simulated slants ranging from 22° to 70°.

Methods:

Subjects. The three subjects in this experiment were graduate students familiar with psychophysical experiments (JB, the author, JL and MC). JL and MC were naive as to the purpose of the experiment.

Design. There were two independent variables, namely the orientation of the planar surface (5 levels: $22, 34, 46, 58$, and 70° slant), and the type of projection (orthographic or perspective).

Stimuli. Each display consisted of two components, a simulated planar patch, and a so-called orientation probe. The patch was rendered by dots distributed on its surface such that there was no texture gradient in the image (when the plane had not been rotated about the vertical). Vertical clipping planes prevented the edges of the plane from becoming visible (as they could serve as 2-D cues). Consequently, dots appeared and disappeared along the vertical edges of the stimulus. The average number of visible dots on the plane was 100. The stimulus was 6.6° wide and 5.6° high. This size was chosen so as to be similar to the stimuli used in the experiments of Chapter 6.

I made the angular velocity of the plane inversely proportional to its slant in an effort to equate the linear image velocities across the different slant conditions. Consider a point $P = (x, y, z)$ rotating about $\Omega = (0, 1, 0)$ through the origin, under orthographic projection. After θ degrees of rotation $P' = (c_\theta x - s_\theta z, y, c_\theta z + s_\theta x)$. The image velocity u at P' is proportional to the depth at P': $u = c_\theta z + s_\theta x$. The total magnitude of the velocity of P during a rotation from 0 to θ_a degrees is then

$$\begin{aligned} U_p &= \int_0^{\theta_a} (c_\theta z + s_\theta x) d\theta \\ &= (s_\theta z - c_\theta x)\Big|_0^{\theta_a} \\ &= s_{\theta_a} z - c_{\theta_a} x + x \\ &= s_{\theta_a} z + (1 - c_{\theta_a}) x. \end{aligned} \tag{5.1}$$

Summing U_p over the right quadrant of the image (i.e., $x \in [0, X]$ and $y \in [0, Y]$) and substituting $z = y \tan \sigma$, we obtain

$$\begin{aligned} U_I &= \int_0^X \int_0^Y (s_{\theta_a} y \tan \sigma + (1 - c_{\theta_a}) x) dx\, dy \\ &= s_{\theta_a} \tan \sigma X Y^2/2 + (1 - c_{\theta_a}) X^2 Y/2. \end{aligned} \tag{5.2}$$

The motion in the other three quadrants is the same if we consider only the magnitude of the velocity and not its direction. Clipping has no effect on this computation because for each point that disappears another one appears, at a different location, having the same relative depth and hence image velocity.

Given a plane slanted by σ_1 and a total rotation of θ_1, Equation 5.2 allows us to compute the amount of rotation θ_2 for a plane slanted by σ_2 that would equate the total motion for both planes. I used the same angular velocity under perspective projection as was computed under orthographic projection.

Although this procedure yields stimuli which have the same total amount of motion, the motion *patterns* for the different slants are quite different. For small slants, motion is largest along the vertical sides of the plane, whereas for large slants motion is largest along the horizontal sides of the planes.

I computed the following angular velocities, θ, using $X = Y = 1°$ and $\theta_a = 2\theta$: $17.17°/s$ (22° slant), $13.15°/s$ (34°), $9.73°/s$ (46°), $6.77°/s$ (58°), and $4.10°/s$ (70°). Each stimulus rotated back and forth, moving for 2 seconds in each direction; the direction of rotation reversed instantaneously. Each frame was shown for 100ms and then immediately replaced by the next one (OSA = 100ms and ISI is zero).

The other component of each display was the so-called orientation probe, a line segment 4° long and positioned well above the rotating plane. Its orientation in the image could be adjusted interactively through a joystick (specifically, the segment rotated about an axis through its midpoint and parallel to the viewing direction; the possible orientations ranged from 0°, horizontal, to 90°, vertical). The advantage of this setup is that the subject can look at the slanted plane and the orientation probe simultaneously; the disadvantage is that the subject has to equate slant in depth with an orientation in the image plane.

Apparatus. Stimuli were displayed on an IMI-455N. Subjects viewed the display in a dark room through a viewing tube which limited the field of view to approximately 10° of visual angle. The distance between the subject's eye and the screen was .9m.

Procedure. Subjects were instructed to equate the orientation of the probe with the perceived orientation of the planar patch. Each session lasted about 30 minutes.

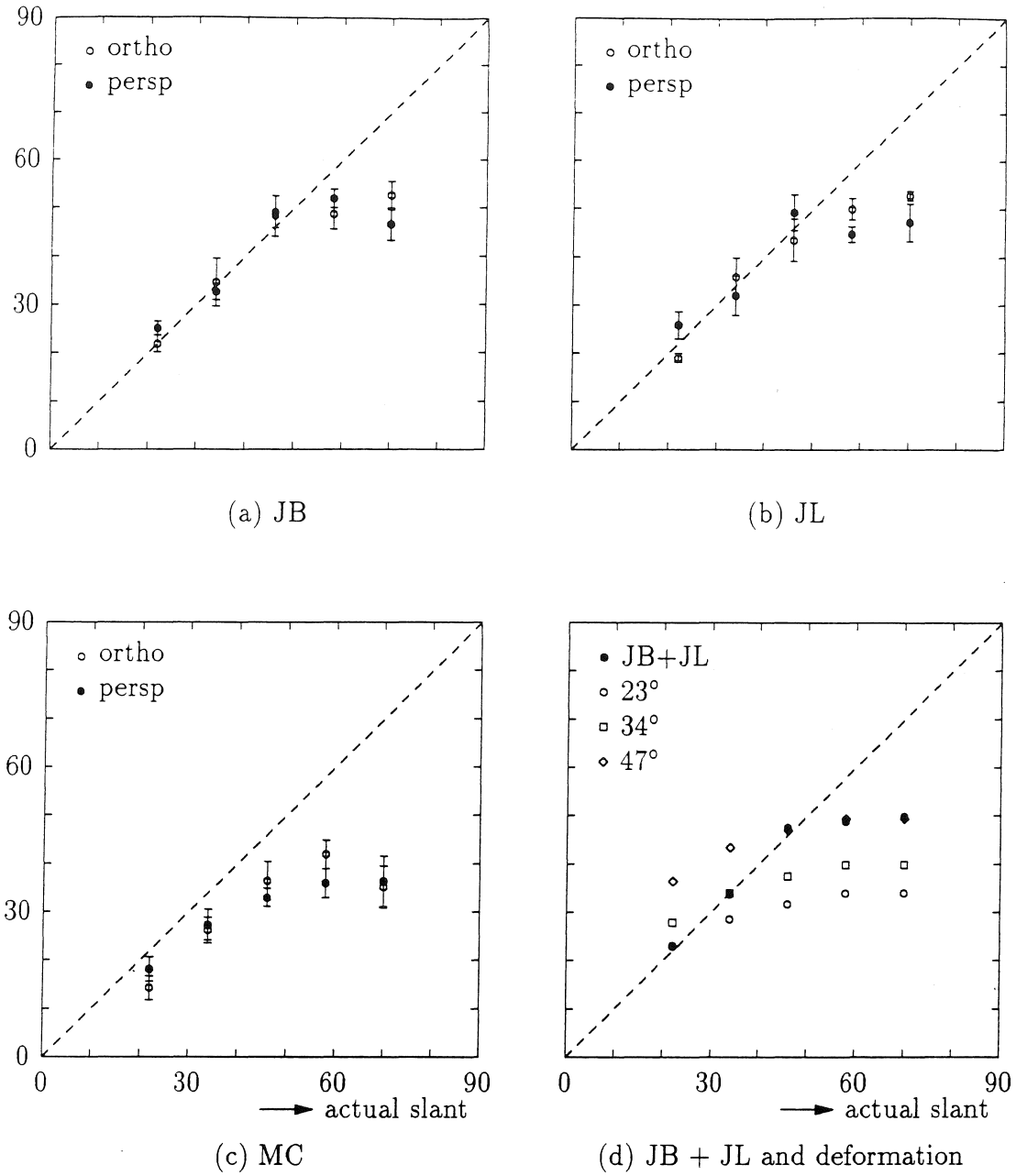
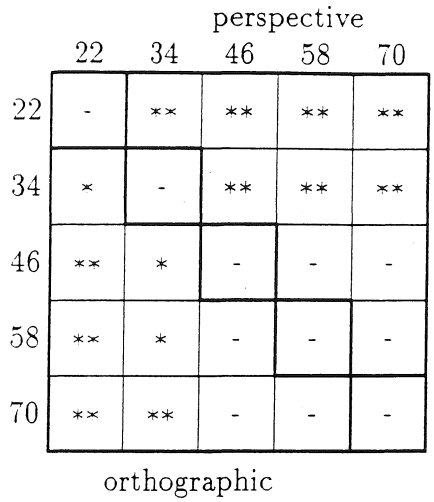
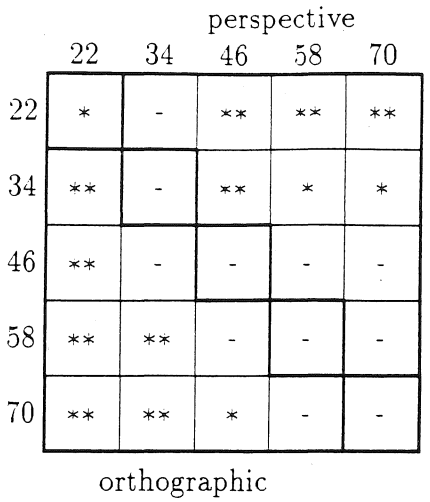


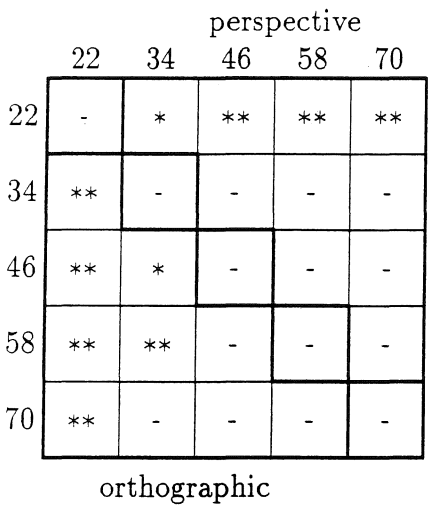
Figure 5.2. Slant judgments for the three subjects (a,b,c); data are means of five trials; error bars indicate standard deviation. (d) Filled circles are combined means from JB and JL, summed over perspective and orthographic projection. Open symbols indicate perceived slant had it been based solely on the deformation of the motion field; circles when equating deformation at 23° slant with 23° slant, squares when anchoring deformation at 34° slant, diamonds when anchoring deformation at 47° slant.



(a) JB



(b) JL



(c) MC

Figure 5.3. Significance of differences in slant judgments shown in Figure 5.2. Entries along the northwest to southeast diagonal refer to orthographic vs. perspective comparison; entries above the diagonal refer to differences between slant judgments under perspective projection; entries below the diagonal refer to differences under orthographic projection. - means not significant at 5% level, * means significant at 5% level, and ** means significant at 1% level.

Results and discussion:

Figure 5.2 shows the judged slant as a function of simulated slant for the three subjects. Figure 5.3 shows how significant the differences between orthographic and perspective projection are, and how significant the differences between the different slant judgments are (for orthographic and perspective projection separately). None of the subjects reported basing his response on some 2D image cue. The following conclusions can be drawn from these results.

(i) There is no significant difference between orthographic and perspective projection (the only significant difference occurs at 22° for subject JL).

(ii) Slant of 22° is distinguishable from all others (except for slants of 22° and 34° under perspective projection for subject JL).

(iii) Slant of 34° is distinguishable from all others for subjects JB and JL (except for slants of 34° and 46° under orthographic projection for subject JL). For subject MC, 34° is only distinguishable from 46° and 58° slant under orthographic projection.

(iv) Simulated slants of 46°, 58°, and 70° are indistinguishable (except for slants of 46° and 70° under orthographic projection for subject JL).

(v) Combining judgments under perspective and orthographic projection for each subject leads to the following unqualified statements: Simulated slants of 22° and 34° are distinguishable from all others (at 1% significance level), and simulated slants of 46°, 58°, and 70° are indistinguishable. This is true for all subjects.

(vi) Judged slants for 22°, 34°, and 46° simulated slants are veridical for subjects JB and JL (see Figure 5.2d for combined judgments of JB and JL), which agrees with the findings of Braunstein, Tittle, & Liter (1990). Subject MC systematically underestimates slants (during debriefing MC mentioned that he has a tendency to see uphill slopes as steeper than they really are, which corresponds to an underestimation of slant).

The question now is why subjects systematically underestimate slants larger than 45°, and in fact judge them to be between 45° and 50°. It cannot be attributed to paucity of dots on the simulated surface (there were about 100 dots) or limited viewing time (subjects could take as long as they felt necessary).

One possibility is that the velocity gradient became to steep for slants larger than 45°. This does not seem to be the case as the following comparison with the experiments in Braunstein & Tittle (1988) argues. The velocity gradient, $\mathbf{u}_y$, in the vertical or y direction for 70° slant was $.29°/s$ per degree visual angle (as measured from the center of the display where image velocity is zero to the top where it is maximal). This gradient falls in the middle of the range of velocity gradients used by Braunstein & Tittle (1988, Experiments 1 and 2). They used dihedral angles translating along the horizontal under perspective projection; for a dihedral angle pointing towards the viewer, image velocity is highest along the horizontal and gradually decreases with distance from the horizontal. For a dihedral angle pointing away, minimal and maximal velocities are reversed. The ratio between maximal ($\mathbf{u}_{max}$) and minimal ($\mathbf{u}_{min}$) image velocity was called the velocity ratio. The loci of maximal and minimal velocity were 6° of visual angle apart. The smallest velocity gradient occurred in their 1.12 velocity ratio condition: $\mathbf{u}_{max} = 6.33°/s, \mathbf{u}_{min} = 5.67°/s$, so that $\mathbf{u}_y = .11/s$. The 2.0 and 3.0 velocity ratios gave rise to the same velocity gradient, namely, $(8°/s - 4°/s)/6° = .67/s$ and $(6°/s - 2°/s)/6° = .67/s$, respectively. Braunstein & Tittle (1988) found that the velocity

Table 5.2 A comparison of the conditions in Experiment 3 of Braunstein & Tittle (1988) and the present study. Top part of the table pertains to Braunstein & Tittle (judged slants from their Figure 6), the bottom part to the present study (judged slants are from Figure 5.2d).

Average Velocity	Angular Velocity	Velocity Ratio	Def	Simulated Slant	Judged Slant
$6°/s$	-	1.12	$.112/s$	47°	28°
$3°/s$	-	1.23	$.112/s$	65°	34°
$6°/s$	-	1.23	$.225/s$	65°	41°
$3°/s$	-	1.52	$.227/s$	77°	45°
-	$17.2°/s$	-	$.121/s$	22°	23°
-	$13.2°/s$	-	$.155/s$	34°	34°
-	$9.7°/s$	-	$.176/s$	46°	47°
-	$6.8°/s$	-	$.192/s$	58°	48°
-	$4.1°/s$	-	$.192/s$	70°	50°

ratio of 1.12 resulted in more accurate judgments of the orientation of the dihedral angle (center near vs. center far) than did the ratios of 2.0 and 3.0. Interestingly, accuracy was the same for ratios of 2.0 and 3.0, which have the same deformation, suggesting that it is perhaps the deformation that determines accuracy. They also found that depth judgments were less accurate for higher velocity ratios (1.57 vs. 1.12), and that depth was more underestimated for the high ratios (which corresponds to underestimating slant).

Table 5.2 shows the stimulus parameters of Experiment 3 of Braunstein & Tittle (1988). I listed the average image velocities, the velocity ratios, the deformations of the projection of the planes making up the dihedral angle, and the simulated and judged slants. I listed the same parameters for the stimuli of the present experiment, except for the velocity ratios which are undefined and the average image velocity, as well as the angular velocities. I computed the deformation of the velocity fields of the translating and rotating planes as described in sections A.12 and A.13, respectively; the axis of contraction in each case is a diagonal, corresponding to a tilt of the slant field of 180° with respect to the vertical (the axis of contraction bisects the direction of tilt and the direction of $\mathbf{A}_t$). The following discussion focuses exclusively on the magnitude of deformation. Judged slant covaries with the deformation in both experiments. Of particular interest is the 65° simulated slant: With an average image velocity of $6°/s$, it is judged to be 41°, and with an average velocity of $3°/s$ it is only 34°. First note that the velocity ratio is the same for both conditions. Second, deformation is smaller for the smaller of the two image velocities, as is judged slant. This suggests

that judged slant is proportional to deformation. To evaluate this suggestion, I converted deformation values into perceived slant angles, σ, according to $\sigma(\text{def}) = \text{atan}(\text{def} \tan \sigma_0/\text{def}_0)$, where def_0 is the magnitude of the deformation associated with some perceived slant angle σ_0. Figure 5.2d shows the results for the rotating planes with $\sigma_0 = 23°, \sigma_0 = 34°$ and $\sigma_0 = 47°$. Clearly, the range of the resulting slant values is too small in each of the three cases. The results are better for the translating planes. With $\sigma_0 = 28°$, $\sigma(.225) = 46.9°$; the perceived slant at the average velocity of σ_0 is 41°. Similarly, with $\sigma_0 = 34°$, $\sigma(.227) = 53.8°$ and the corresponding perceived slant is 45°.

I should point out that it is actually easy to compute the correct slant for the translating planes that make up the dihedral angle. As reviewed in section A.6, $\sigma = \text{atan}(\text{def}/\text{A}_\text{t})$; for planes translating in a frontoparallel direction, A_t is simply the magnitude local image velocity in radians/s. For example, the simulated slant of 65° with an average image velocity of $3°/s$ yields $\sigma = \text{atan}(.122 \times 180/\pi/3) = \text{atan}(2.14) = 64.9°$. Why, then, does the human visual system not use this readily available information? It might be because it is lost as soon as the eyes perform a smooth pursuit eye movement to track the moving dihedral angle (it could be made available through the angular velocity of the eyes); only differential invariants such as the deformation remain available. Thus, it is not unreasonable that the visual system does not take average image velocity into account. Nevertheless, there is a small effect of average image velocity. For example, a deformation of $.112/s$ in the context of an average image velocity of $3°/s$ leads to a judged slant of 34°; the same deformation with an image velocity of $6°/s$ leads to 28°. Thus, the same deformation leads to slightly higher slant judgments with the smaller of the two image velocities, as would be expected.

For the rotating planes the situation is more difficult because $A_t = \omega_z$, the angular velocity of the planes. The results of experiments in section 6.2, in which subjects have to equate the angular velocity of two objects rotating about the vertical, suggest that angular velocity is hard to judge and that it requires an effort not to simply equate linear image velocities. Note that the average image velocity for the rotating planes cannot be used to compute slant. Linear image velocity varies greatly across the image, as it is proportional to the local relative depth, that is, depth with respect to the axis of rotation. Close to the center of the plane velocities are small, leading to slant angles close to 90°. On the other hand, it remains the case that slant judgments were veridical between 23° and 47°, suggesting that angular velocity is perceived correctly and used to convert deformation into slant. Note that smooth pursuit eye movements are not a problem in this situation (assuming they would even occur). Any point on the plane can be foveated and the correct angular velocity of the plane can, in theory, still be recovered. Assuming that angular velocity is computed and used for simulated slants of less than 47°, why is it not computed correctly for the larger slants?

To summarize, at least under the conditions of the present experiment and those of Braunstein & Tittle (1988), the visual system seems to respond largely on the basis of deformation *per se* without taking A_t into account. The range of slant judgments is comparable in both experiments—28° to 45° for translating planes, versus 23° to 50° for rotating planes—as is the range of deformations—.113/s to .228/s versus .121/s to .197/s. What appeared to be the simpler situation, namely an observer looking at a translating plane as opposed to a rotating one, is actually the one that is least informative in the context of the human visual

system which engages continuously in smooth pursuit eye movements.

Other, more prosaic, causes for the underestimation of slant need to be kept in mind. Accommodation, the constant brightness of the dots, and the absence of size changes in the dots all indicate that the object of observation is in fact a frontoparallel plane without any slant. The pattern of slant judgments, however, argues against such nonspecific effects. The slants of the translating and rotating planes simulating a slant of about 70° were both underestimated by about 20°. This could be attributed to the influence of the flatness cues mentioned above. But if this were the reason, one would expect other simulated slants to be underestimated in the same manner for translation and rotation. This is not the case. A simulated slant of 47° is judged accurately under rotation but is underestimated by 20° under translation.

One reason for studying how human observers perceive and represent the orientation of a planar surface patch is that one can think of a solid object as 3D space bounded by small, planar surface patches. One assumption is that the surface surrounding a patch does not influence its perceived orientation. This is probably true as a first approximation, but it is worth keeping in mind that it is an assumption that proved to be wrong in the case of brightness perception (Cornsweet, 1970). Moreover, Rogers & Graham (1983) found an illusion similar to the Craik-O'Brien-Cornsweet illusion in the luminance domain for perceived depth in stereo and motion parallax. And Stevens & Brookes (1987a, 1988) argue that perceived depth is constructed by integrating between those parts of a surface that are curved or that exhibit a depth discontinuity much as brightness is constructed from discontinuities in luminance. Perhaps changes and discontinuities in deformation determine perceived slant rather than deformation itself.

To summarize, the underestimation of slant of rotating planar patches under orthographic projection is not due to the lack of perspective information. It is, however, correlated with deformation, one of the first-order differential invariants of the image velocity field (Koenderink & van Doorn, 1975, 1976b, 1986; Koenderink, 1986; Rogers & Koenderink, 1986, used deformation in their explanation of Ogle's induced effect). The results of the present experiment and of Braunstein & Tittle (1988) can be explained semi-quantitatively if we assume that judged slant is proportional to deformation.

Chapter 6
Visual perception of solid shape from occluding contours

In this chapter, I will investigate whether human observers perceive solid shape when looking at the occluding contours of a rotating solid object, in the present experiments, a rotating ellipsoid. In Experiment 6.1, subjects have to indicate whether they perceive a solid, volume-containing ellipsoid or a flat, planar ellipse while looking at their respective occluding contours; performance is a measure of how well occluding contours lead to a percept of solid shape. According to the models developed in Chapter 4, an observer first computes an object's angular velocity and then its shape. In Experiment 6.2, I study how well observers can perceive an object's angular velocity. In the surface interpolation Experiment 6.3, subjects specify perceived shape by adjusting the depth of a probe point such that it appears to lie on the surface of the perceived solid object.

I used orthographic projection throughout the following computer simulations and psychophysical experiments (see Chapter 5 for a discussion of perspective versus orthographic projection).

6.1 Discriminating between solid and planar objects

In section 4.3, I formulated a number of models for the perception of solid shape from the occluding contours of a rotating object. The purpose of the experiments described in this section is to determine which, if any, of these models apply to the human perception of shape from contours.

The subject's task in Experiment 6.1 is to discriminate between planar ellipses and solid ellipsoids; that is, the subject looks at the contours of either a rotating planar ellipse or a rotating solid ellipsoid and has to decide which of the two it is. The specific instructions are to indicate whether the perceived object is flat (a planar ellipse) or not (a solid ellipsoid or a nonrigid object).

The design of this experiment is based on the fact that any rim of an ellipsoid, that is, the pre-image of any occluding contour, is itself a planar ellipse; this is true under orthographic as well as perspective projection (section A.4). It is impossible, therefore, to discriminate between the contours of an ellipsoid and those of an ellipse on the basis of their shape (both are elliptic) or their instantaneous velocity field (both derive from a rotating planar ellipse; see section 4.2, where it is shown that the instantaneous velocity field along an occluding contour is determined solely by the corresponding rim, i.e., it does not depend on the curvature of the surface in the neighborhood of the rim). Thus, the subject is forced to integrate more than two views into his percept in order to perform the task (assuming, of course, that there are no trivial 2D cues to make the discrimination).

By viewing a planar ellipse as an ellipsoid which happens to be infinitesimally thin along one of its axes, we can create a one-parameter family of ellipsoids which includes the ellipse

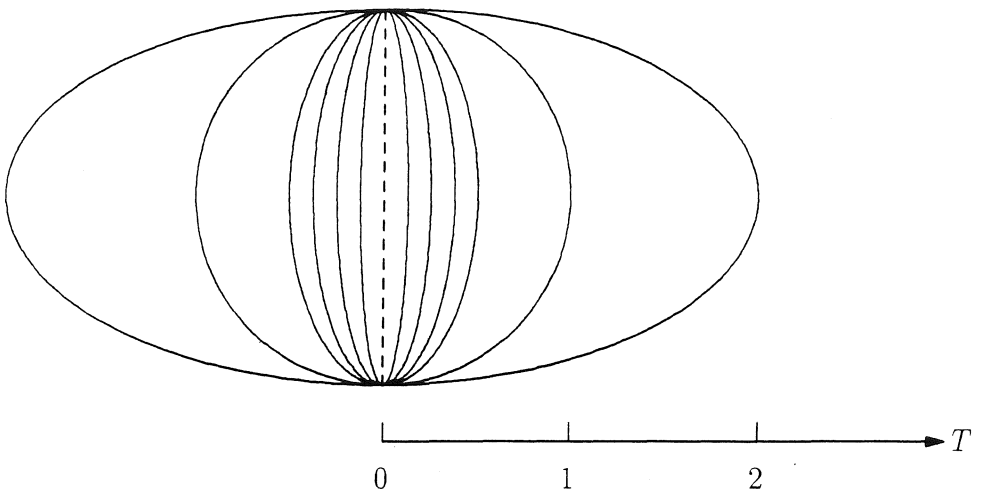


Figure 6.1 A side view of a one-parameter family of ellipsoids. T is the thickness of the ellipsoids in the horizontal direction. The planar ellipse, depicted as a dashed line and for which $T = 0$, is the extreme member of this family.

as its extreme member. The parameter will be referred to as the "thickness" T of the ellipsoids. Figure 6.1 shows a side view of a family of ellipsoids whose thickness varies from 0 to 2. The idea is to determine how well a subject can discriminate between an ellipse and members of its associated family of ellipsoids. Thus we determine the psychometric function $p(f(T))$, specifying the probability that a subject will respond with "fat ellipsoid" as a function of $f(T)$, a function of the ellipsoid's thickness. For $f(T)$, we will use the following discrimination measures derived from the different models for the perception of solid shape from occluding contours.

Curvature model. Consider an ellipsoid rotating about a vertical axis. $R_m(s,t)$ is the radius of curvature of a horizontal cross section at point s on the ellipsoid's rim at time t (Figure 4.9). As $R_m = 0$ everywhere for a rotating ellipse, and as R_m increases with the thickness of an ellipsoid, a plausible measure of discriminability associated with the curvature model is

$$C_{R_m} = \int\int R_m \, ds \, dt \, . \tag{6.1}$$

If, instead of R_m, surface curvature along the rim is specified by the inverse of radial, Gaussian, or mean curvature we get the corresponding measures of discriminability, namely, C_{R_r}, C_G, and C_H, respectively.

Rim prediction model. The instantaneous velocity field along the contour of a rotating object gives relative depth along its rim (i.e., depth relative to the axis of rotation). This

information is used to predict subsequent rims and hence contours, given the current estimate of the angular velocity of the ellipsoid. Suppose we are looking at a rotating ellipse, and assume we know the correct angular velocity, then the predicted rim will be the actual rim; intuitively, radial curvature along the rim of an ellipse is infinitely large, making it impossible for the rim to slide over the surface of the ellipse. The fatter the ellipsoid, the smaller the radial curvature along its current rim and the easier the rim can slide over its surface. Thus, the fatter the ellipsoid, the more the predicted rims will differ from the actual rims.

Let $P_t(s)$ be the coordinates of the point s on the rim as determined by the observer at time t; let $P_t^f(s)$ indicate the predicted appearance of the rim at some future time f, based on the rim observed at time t. For an ellipse, $P_f = P_t^f$ for all t, that is, predicted and actual rims are always identical. The larger the difference between P_f and P_t^f, the fatter the ellipsoid. This suggest the following discrimination measure associated with the rim prediction model

$$\mathcal{R} = \int_{t=0}^{T} \int_{f=t}^{f=t+\tau} (P_f - P_t^f) df\, dt, \tag{6.2}$$

where τ is the period of time during which predictions based on a particular rim are compared with actual rims. If $\tau = 0$ the predicted rim is only compared with itself, so that $\mathcal{R} = 0$. In the other extreme case, $\tau = T$, so that each predicted rim is compared with all the rims that follow it. For an ellipse, $\mathcal{R} = 0$ regardless of τ; for an ellipsoid, $\mathcal{R}$ increases with τ and with $P_f - P_t^f$.

Nonrigidity model. Although not explicitly based on a model of 3D shape perception, the nonrigidity measure is of interest because it is related to the preceding one and because it correlated well with the outcome of psychophysical experiments on the discrimination of rigid and nonrigid motion (Braunstein, Hoffman, & Pollick, 1990). In their experiments, subjects look at a small number of moving dots and have to decide whether they form a rigid, 3D structure or not. In rigid motion, the 3D distances among the points do not change, whereas they do in nonrigid motion. As a measure of nonrigidity, call it $\mathcal{N}$, Braunstein et al. proposed the mean of the variance in interpoint distances. That is, pick any two points and compute the variance in their 3D distance over time; compute this variance for all possible pairs of points; the mean of these variances is the measure of nonrigidity. For points that are moving rigidly, $\mathcal{N} = 0$ because the variance in all interpoint distances is by definition zero. The d' in this experiment was proportional to $\mathcal{N}$.

Because the rim of a rotating ellipse does not change over time (only its position in space does), it can be considered as a collection of points that are rigidly connected. The rim of a rotating ellipsoid does change so that it can be viewed as a nonrigid configuration of points. Thus, the distinction between flat ellipses and fat ellipsoids could be based, not on shape, but on the presence or absence of rigid motion.

Tangent plane prediction model. This model is an elaboration upon the rim prediction model; in addition to depth along the rim, we now also explicitly include the orientation of the local tangent plane in the percept of shape.

Let $T_t(s)$ specify the tangent plane at a point s on the rim at time t. By definition, $T_t(s)$ contains the local viewing direction for all s. Let $T_t^f(s)$ be the predicted tangent plane

along the predicted rim at time f, based on the rim and tangent plane observed at time t. For a rotating ellipse, T_f will never equal T_t^f, that is, predicted and actual tangent planes will never be the same, because T_f includes the current viewing direction whereas T_t^f does not (unless the object rotated a multiple of 180°); this is true for ellipse as well as ellipsoids. The fact that $T_f \neq T_t^f$ whereas $P_f = P_t^f$ indicates that surface curvature along the rim of an ellipse is infinitely large; in other words, a particular location on the object's surface has different tangent planes, which is only possible if surface curvature is infinite at that location. For a rotating ellipsoid, the predicted and actual rims never coincide, so this situation never arises. If we measure the change in tangent plane between the predicted and actual rim as a function of their distance, the rotating ellipsoid and ellipse can be treated in the same way. Thus, the measure of discriminability associated with the tangent plane prediction model becomes:

$$\mathcal{T} = \int_{t=0}^{T} \int_{f=t}^{f=t+\tau} \frac{P_f - P_t^f}{T_f - T_t^f} df\, dt, \tag{6.3}$$

Note that $\mathcal{T} = 0$ for the ellipse because $P_f - P_t^f = 0$ and $T_f - T_t^f \neq 0$. Intuitively, the fatter the ellipsoid the smaller the change in tangent plane per unit distance along its surface, that is, the larger $\mathcal{T}$ will be.

I computed the above discrimination measures for the one-parameter family of ellipsoids

$$\mathcal{E}_T : x^2 + y^2/T + z^2/6 = 1, \tag{6.4}$$

rotated 30° about the x-axis (in other words, the angle between the ellipsoids' major axis and the vertical is 60°). Figure 6.2 shows five views of four members of this family as they rotate about a vertical axis; the amount of rotation between views is 15°; I used orthographic projection. The top panel shows views of the flattest ellipsoid, $T = 0$, the bottom panel of the fattest ellipsoid, $T = 1.18$. It is clear from these sequences that the vertical size of the contours increases with T, as does their area in the image. Also, side views of the flatter ellipsoids are more eccentric. These are some of the 2-D cues that covary with the thickness of an ellipsoid; I will return to them in Experiment 6.1.

Contours are approximated by 120 straight line segments, which is sufficient to gain a smooth appearance. The endpoints of the line segments are equally spaced in the vertical direction. Let $P(i,\theta) = (P_x, P_y, P_z)$ be the coordinates of the i-th endpoint after rotating the ellipsoid by θ°; x is in the horizontal direction, y is in the vertical direction (parallel to the axis of rotation); z is distance relative to the axis of rotation. $p(i,\theta)$ is the projection of $P(i,\theta)$, that is, $p(i,\theta) = (P_x, P_y)$. $dp(i,\theta)$ denotes the distance between $p(i,\theta)$ and $p(i-1,\theta)$, and $ds(i,\theta)$ is $(dp(i+1,\theta) + dp(i,\theta))/2$. Finally, let $\triangle\theta$ be the amount of rotation between successive views.

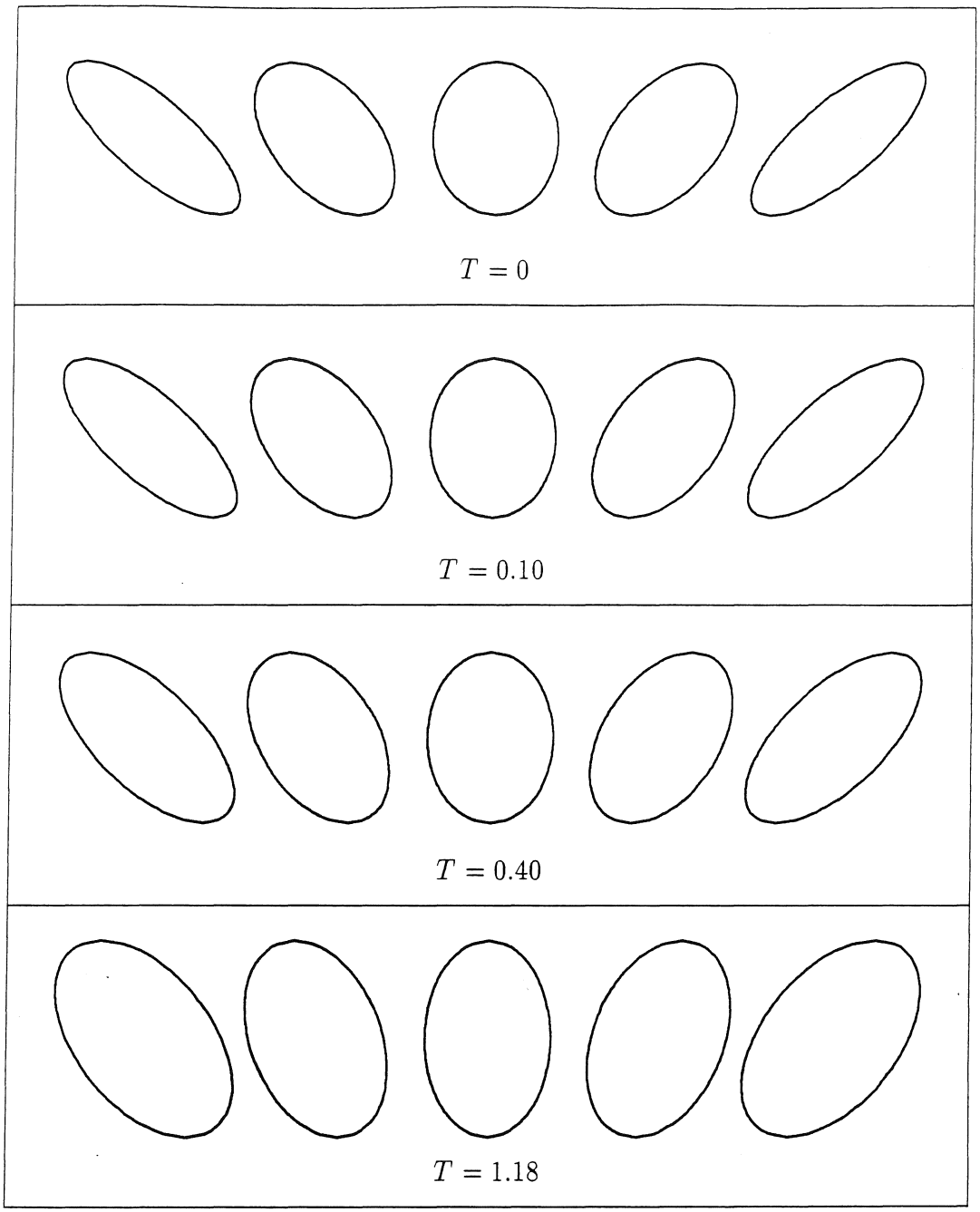


Figure 6.2. Five views each of four members of the family of ellipsoids $x^2 + y^2/T + z^2/6 = 1$, rotated 30° about the x-axis. Ellipsoids rotate about the vertical; 15° rotation between views. Contours have been displaced horizontally for clarity.

The radius of curvature, $R_m(i,\theta)$, is computed as

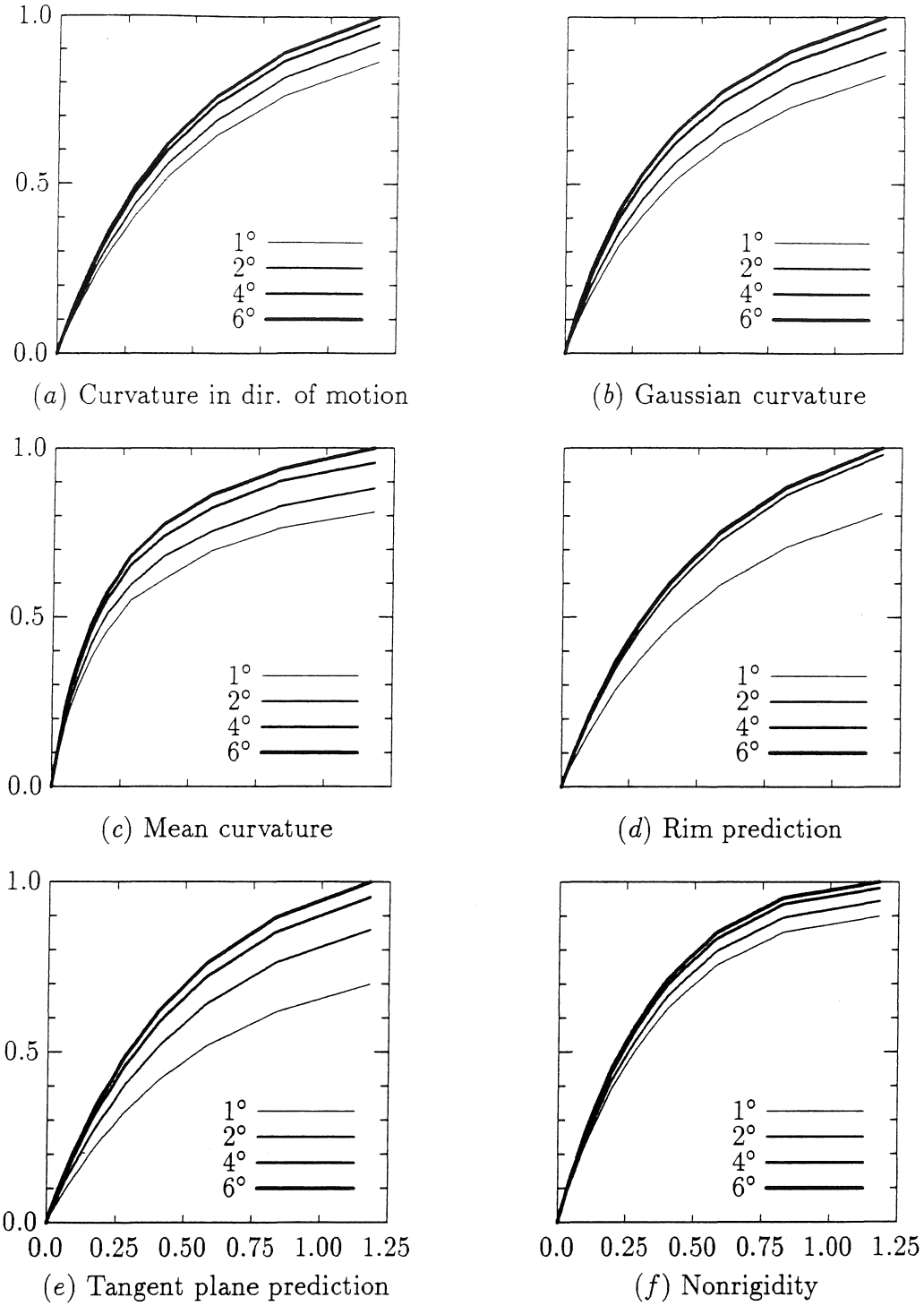
$$R_m(i,\theta) = P_x(i,\theta) + [(P_x(i+1,\theta) - P_x(i,\theta)) - (P_x(i,\theta) - P_x(i-1,\theta))]/(\triangle\theta)^2, \tag{6.5}$$

and the corresponding discrimination measure is $\mathcal{C}_{R_m} = \sum_\theta \sum_i R_m(i,\theta) ds(i,\theta) \triangle\theta$. The reason for weighting the local measurements R_m by ds, that is, a weight derived from the 2-D contour instead of the 3-D rim, is that the 3-D location of the rim is assumed to be unknown in the curvature model. Anyway, it would make little difference if, as we will do in the sequel, one considers a family of ellipsoids whose rims are all slanted by more or less the same amount. Figure 6.3a shows $\mathcal{C}_{R_m}$ for the family of ellipsoids $\mathcal{E}_T$ with $0 \leq T \leq 1.18$, and for four different values of $\triangle\theta$, namely, 1° (thickest line), 2°, 4°, and 6° (thinnest line). As expected, $\mathcal{C}_{R_m}$ is zero for an elliptic disc ($T = 0$) and increases with the ellipsoid's thickness. Also, $\mathcal{C}_{R_m}$ increases as $\triangle\theta$ gets smaller; this effect levels off for $\triangle\theta \leq 2°$.

Radial curvature, κ_r, at a point on the rim is computed from $\kappa_r = \cos(\alpha)\kappa_m$, where $\kappa_m = 1/R_m$, and α is the angle between the local normal to the contour and the horizontal (i.e., the direction of motion). I computed the transverse curvature, κ_t, that is, the curvature of the contour itself, from the exact expression for the contour (which is an ellipse). Gaussian curvature, K, is then obtained from $K = \kappa_r \kappa_t$ (Koenderink, 1984b). I computed mean curvature, H, from the formula $H = \kappa_t(1 + \kappa_r/\kappa_{rim})/2$, where κ_{rim} is the normal curvature of the ellipsoid sectioned in the direction of the rim (Giblin & Weiss, 1987). The corresponding discrimination measures are $\mathcal{C}_K = \sum_\theta \sum_i ds(i,\theta)/K(i,\theta)\ \triangle\theta$ and $\mathcal{C}_H = \sum_\theta \sum_i ds(i,\theta)/H(i,\theta)\ \triangle\theta$. They are shown in Figures 6.3b and c, respectively, for the ellipsoids $\mathcal{E}_T$. Both measures are zero for the flat ellipsoid, $T = 0$, and increase with the ellipsoids' thickness. $\mathcal{C}_H$ increases somewhat faster than $\mathcal{C}_K$ for moderately flat ellipsoids. Again there is a small effect of $\triangle\theta$.

In the rim and tangent plane prediction models, we start by measuring relative depth along the contour: $z(i,\theta) = \big(p_x(i,\theta+\triangle\theta) - p_x(i,\theta-\triangle\theta)\big)/2$. Assuming the angular velocity of the ellipsoid is known, future locations of the rim can be computed and the predicted and actual input compared. The comparison can be between the predicted and actual rims, or between the predicted and actual contours. The corresponding discrimination measures will be referred to as $\mathcal{R}_r$ and $\mathcal{R}_c$. The latter only measures the difference in x-coordinates between predicted and actual contours; the y-coordinates remain the same because the ellipsoid is rotating about a vertical axis. $\mathcal{R}_r$ measures differences in the z- as well as the x-coordinates. For the family of ellipsoids $\mathcal{E}_T$, the value of $\mathcal{R}_c$ is only about 10% of that of $\mathcal{R}_r$; and in case of noisy images, $\mathcal{R}_c$ increases very little with T whereas $\mathcal{R}_r$ continues to increase considerably. Therefore, I will focus exclusively on $\mathcal{R}_r$ in the following.

There is one free parameter that remains to be determined before we can actually compute $\mathcal{R}_r$ and that is the time period τ during which predictions derived from a particular rim are compared with actual rims. This period can be specified in terms of actual elapsed time or the amount of rotation of the object. There does not appear to be any compelling reason to chose one over the other, nor are there empirical data that point to either one. With the understanding that, at this point, my choice is essentially arbitrary, I chose to specify the period of comparison in terms of the amount of rotation. Figure 6.3d shows $\mathcal{R}_r^{20}$, that is, rims are rotated up to 20° to generate predictions. As with the other measures, $\mathcal{P}_r$ increases



(*a*) Curvature in dir. of motion

(*b*) Gaussian curvature

(*c*) Mean curvature

(*d*) Rim prediction

(*e*) Tangent plane prediction

(*f*) Nonrigidity

Figure 6.3. Discrimination measures for the one-parameter family of ellipsoids $\mathcal{E}_T$ as a function of thickness T. Total amount of rotation about the vertical is 60°. Each measure is computed for four different angles between successive views (6°, 4°, 2°, and 1°). Ordinate gives the value of each measure as a fraction of its maximal value.

with the ellipsoids' thickness. There is also an effect of $\triangle\theta$ up to $\triangle\theta = 2°$: for smaller $\triangle\theta$, $\mathcal{R}_r$ is not affected. Figure 6.6a (left panel) shows the expected effect of θ on $\mathcal{R}_r$: the larger τ, the faster and the more $\mathcal{R}_r$ increases.

The tangent plane prediction measure, $\mathcal{R}_T^{20}$, is shown in Figure 6.3e. Of all the measures, $\mathcal{R}_T^{20}$ has the largest effect of $\triangle\theta$, especially in going from $\triangle\theta = 6°$ to $\triangle\theta = 2°$.

Finally, Figure 6.3f shows the nonrigidity measure $\mathcal{N}$. $\mathcal{N}$ is the mean of the variance of all interpoint distances taken between the 120 points that make up the contour.

To summarize, the behavior of the six discrimination measures is very similar for the family of ellipsoids $\mathcal{E}_T : x^2 + y^2/T + z^2/6 = 1$, rotated 30° about the x-axis (slant of the ellipsoids' major axis is 60°), and rotating about a vertical axis. Their behavior remains similar if the ellipsoids of this family are rotated by 60° or 90° about the x-axis (slant of the ellipsoids' major axis is 30° and 0°), although the actual values are considerably smaller. It also remains true for ellipsoids with different shapes.

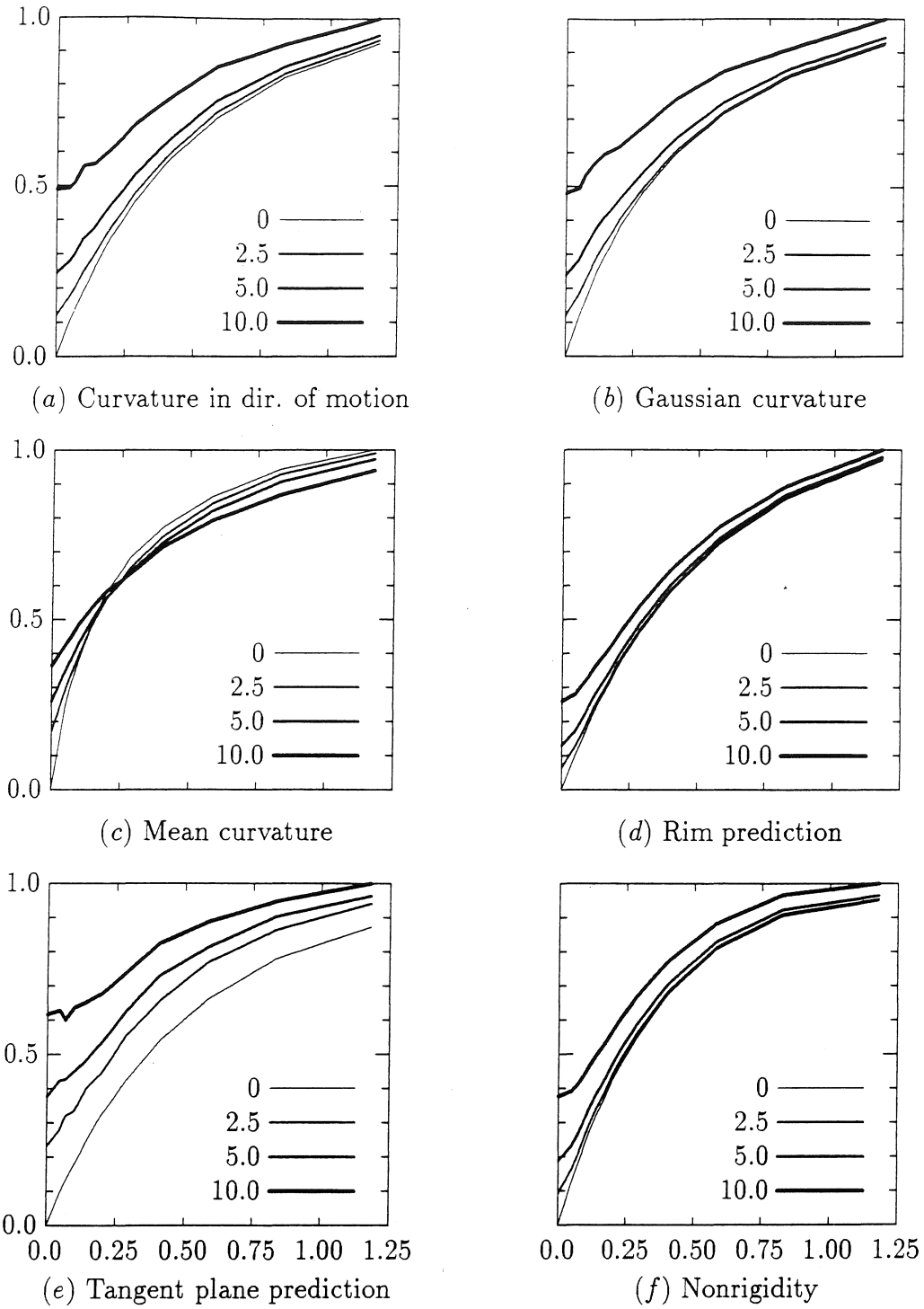
So far I assumed that the position, the velocity and the curvature at points on the occluding contour are known accurately; I also assumed that the angular velocities of the ellipsoids are known. This is clearly not realistic if these simulations are to pertain to physical and hence noisy systems. It is therefore important to determine how the different discrimination measures are influenced by noise in the image and noise in their computations. In the following, I will indicate how I perturbed the computations of each discrimination measure. I perturb a value by multiplying it by the factor $1 + \text{random}[-r, r]$, where $\text{random}[-r, r]$ is a random number between $-r$ and r; and I refer to it as $r \times 100\%$ noise.

The radius of curvature in the direction of motion is

$$\begin{aligned} R_m(i,\theta) &= P_x(i,\theta) + [(P_x(i+1,\theta) - P_x(i,\theta)) - (P_x(i,\theta) - P_x(i-1,\theta))]/(\triangle\theta)^2 \\ &= P_x(i,\theta) + (v(i+1,\theta) - v(i,\theta))/\triangle\theta, \end{aligned} \tag{6.6}$$

where v stands for velocity. I independently perturb $P_x, v(i), v(i+1)$, and $\triangle\theta$. Radial curvature is given by $\kappa_r = \cos(\alpha)\kappa_m$, where $\kappa_m = 1/R_m$, and α is the angle between the local normal to the contour and the horizontal. $1/\kappa_r$ is perturbed by perturbing $\cos\alpha$ and $1/\kappa_m$. The inverse of Gaussian curvature is perturbed by independently perturbing $1/\kappa_r$ and $1/\kappa_t$. The inverse of the mean curvature is perturbed by perturbing $1/\kappa_r, 1/\kappa_t$, and $1/\kappa_{rim}$. The rim prediction measure is perturbed by perturbing the estimated depth along the rim, the location of the current rim, and the amount of rotation used to predict subsequent rims. The tangent prediction measure is perturbed by perturbing the current tangent plane in addition to the perturbations listed under the rim prediction; it was necessary to threshold the difference dT between the predicted and actual tangent planes: Only differences larger than .002 were included (lower thresholds led to wild fluctuations in $\mathcal{T}$, due to the fact that dT occurs as a divisor). The nonrigidity measure is perturbed by perturbing each interpoint distance, which is done by perturbing the difference in x-, y-, and z-coordinates of the two points for which the distance is to be determined, and then perturbing their computed distance.

The resulting noisy discrimination measures are shown in Figures 6.4 and 6.5 for various amounts of noise (I used $\triangle\theta = 2°$ throughout). Figure 6.4 gives the value of each measure as a fraction of its maximal value for 0%, 2.5%, 5% and 10% noise. The measures increase

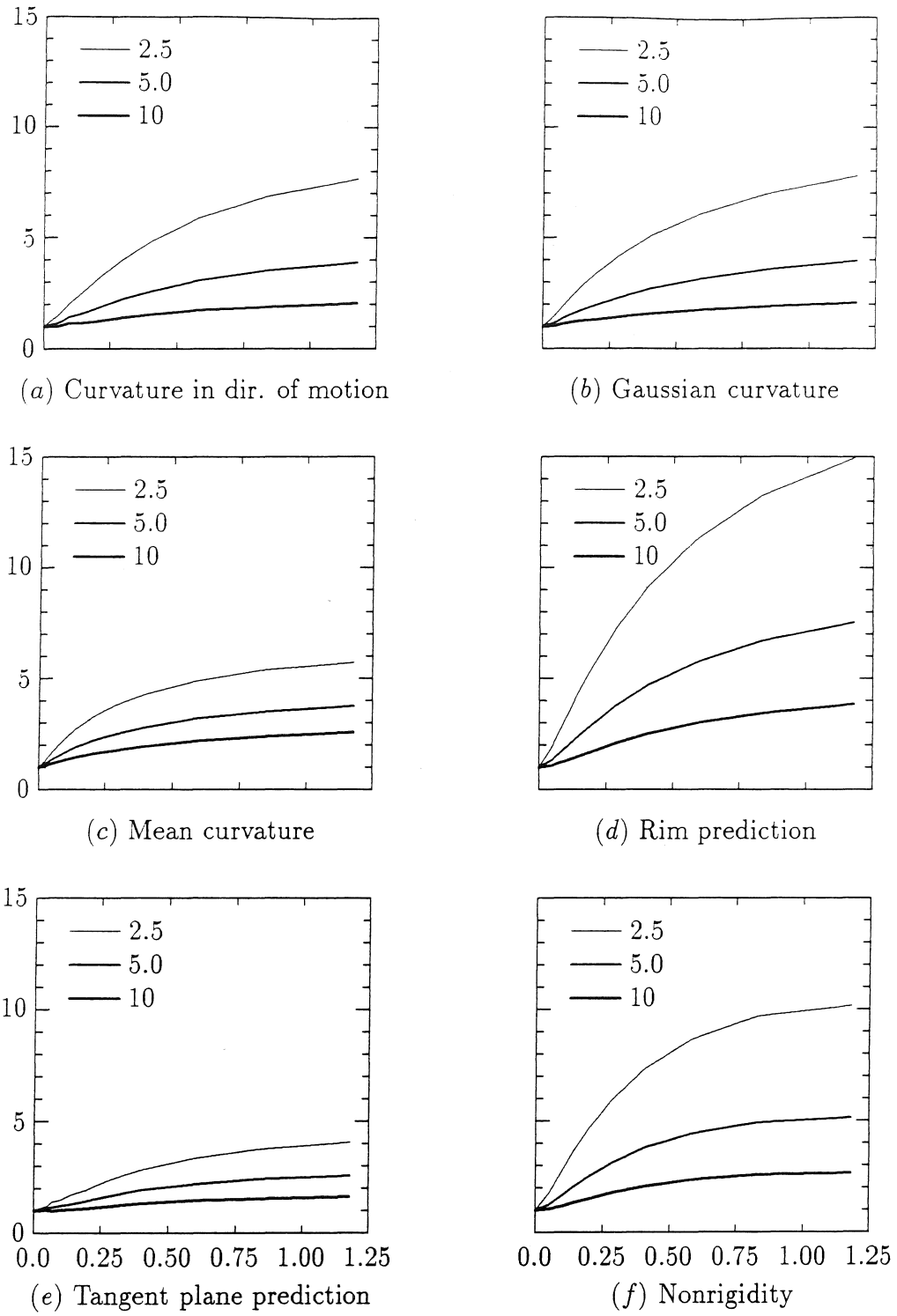


(*a*) Curvature in dir. of motion (*b*) Gaussian curvature

(*c*) Mean curvature (*d*) Rim prediction

(*e*) Tangent plane prediction (*f*) Nonrigidity

Figure 6.4. Effect of noise on discrimination measures for the one-parameter family of ellipsoids $\mathcal{E}_T$. There are four noise levels: 0%, 2.5%, 5.0%, and 10.0%. Abscissa indicates the value of T. Ordinate gives the value of each measure as a fraction of its maximal value. Total amount of rotation about the vertical is 60° about the vertical; angle between views is 2°.



(*a*) Curvature in dir. of motion

(*b*) Gaussian curvature

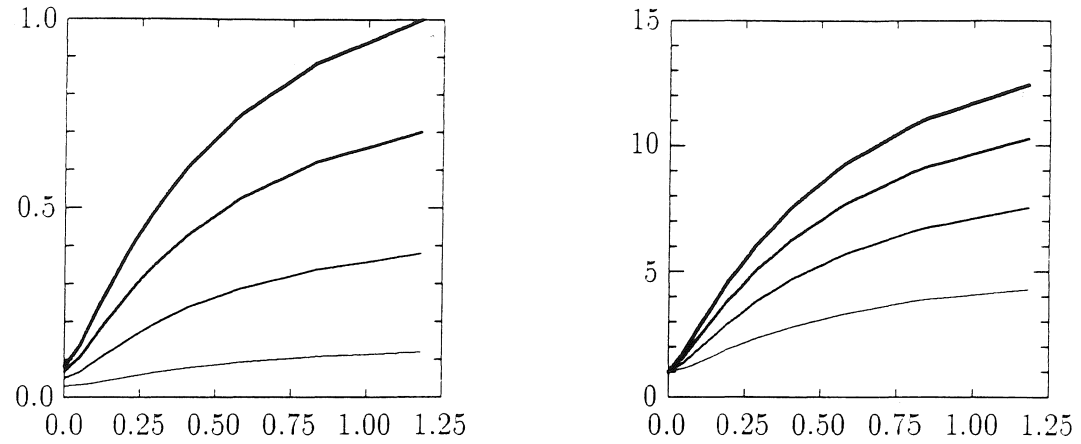
(*c*) Mean curvature

(*d*) Rim prediction

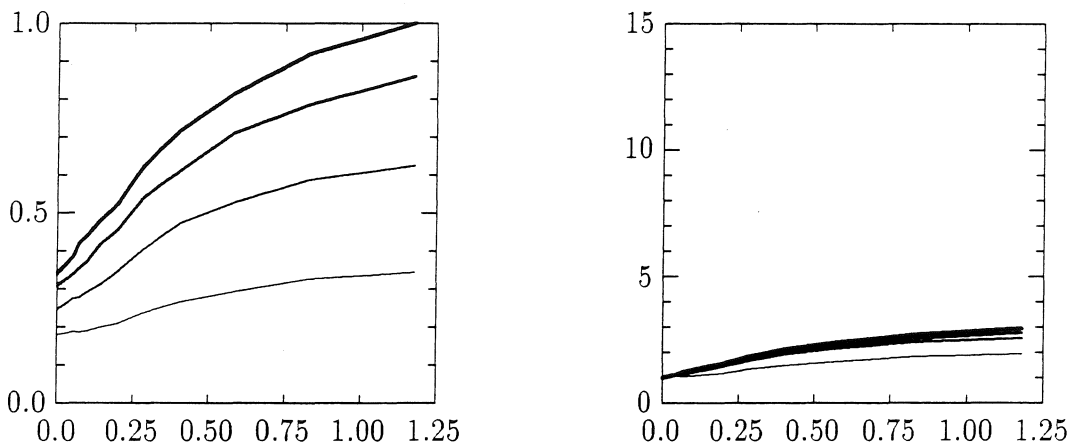
(*e*) Tangent plane prediction

(*f*) Nonrigidity

Figure 6.5. Effect of noise on the increase of discrimination measures relative to their value for the ellipse. There are three noise levels: 2.5%, 5.0%, and 10.0%. Same ellipsoids and data as in Figure 6.3.



(*a*) Rim prediction



(*b*) Tangent plane prediction

Figure 6.6. Effect of the amount of rotation, τ, during which predictions are compared with the input on (a) rim and (b) tangent plane prediction measures. Amount of rotation varies from 10° (thin lines), to 20°, 30°, and 40° (thick lines). Noise is 5% throughout; ellipsoids as in Figure 6.4.
left panels: discrimination measures as fractions of their maximal value.
right panels: discrimination measures as multiples of their value for the ellipse.

considerably with noise, in particular for the flatter ellipsoids. The rim prediction and nonrigidity measures are affected the least. Figure 6.5 gives the value of each measure as a multiple of its value for the ellipse (i.e., $T = 0$), for 2.5%, 5%, and 10% noise. The rim prediction measure is the most robust one, followed by the nonrigidity measure. The least robust is the tangent plane prediction measure.

The performance of the rim prediction measure can even be further improved by increasing τ, the time during which predicted and actual rims are compared. This is illustrated in Figure 6.6a where $\tau = \{10°, 20°, 30°, 40°\}$ and noise is at the 5% level. Interestingly, the tangent plane prediction measure improves hardly at all with increasing τ (Figure 6.6b).

Experiment 6.1: Discriminating between ellipsoids and ellipses

The purpose of this experiment is to determine whether human observers perceive solid shape from dynamic occluding contours and to find evidence for or against a particular computation of shape.

Methods:

Subjects. The author (JB) was the subject in experiments B-1 through B-10. In experiment B-15, JB and JT were subjects.

Design. There was one independent variable, namely the thickness of the ellipsoid (in experiment B-15, the slant of the axis of rotation was a second variable).

Stimuli. Each stimulus consisted of n orthographic views $(v_1, v_2, ..., v_n)$ of an ellipsoid rotating about a vertical axis at constant angular velocity θ. Each view was shown for 100ms and was then replaced immediately by the next one (OSA is 100ms, ISI is zero). Subjects saw a sequence of contours $v_1 v_2 ... v_{n-1} v_n v_n v_{n-1} ... v_2 v_1 v_1 v_2 ...$, corresponding to an ellipsoid rotating back and forth, until they responded. The direction of rotation reversed instantaneously at v_1 and v_n in the sequence. The range of rotation about the vertical is given in column R_V of Tables B.1 through B.12. Contours were computed assuming orthographic projection. In "contour sequences" the occluding contour was drawn in white against a black background; in "silhouette sequences," the background was filled-in with white horizontal lines which were sufficiently close to give an impression of a uniformly white background.

Figure 6.2 shows that as the thickness of the ellipsoid increases so does its vertical size and the area of its projection. To eliminate the vertical size confound, ellipsoids were scaled such that their projections all had the same vertical size, namely 4°. To eliminate the area confound, I varied a_{11}, that is, the ellipsoid's extent in the horizontal direction; the columns labeled **Area** and **Circf** in Tables B.1 through B.8, in Appendix B, show that area and circumference of the projection are not correlated with the ellipsoids' thickness. The scaling and variation of a_{11} also ensured that the horizontal extent of the projected contours remained constant (see columns **Max Ext**). Another confound clearly visible in Figure 6.2 is that the horizontal extent of flat ellipses changes more than that of the fat ellipsoids during the course of the sequence. In the extreme case of a side view, the projection of an ellipse will be a mere line in the image, something which is impossible for a fat ellipsoid. I have been unable to remove this 2D confound, which I call contraction, from the stimuli (see columns **Contr**). Increasing $1/a_{22} = T$ helps somewhat, as does changing the amount of initial rotation (R_V) about the vertical axis of rotation (the fewer side views a sequence of a flat ellipse has, the smaller its contraction will be). However none of these parameters can be changed sufficiently without in turn introducing new confounds (in this case of horizontal

size or of different views of the rotating object). In any event, the results of the experiment suggest that contraction does not determine the outcome.

During pilot experiments in which I used the same angular velocity for all ellipsoids, I noted that the fatter the ellipsoids the slower they appear to be rotating. To counteract this tendency, I increase the angular velocity with the ellipsoids' thickness. The algorithm I used to compute angular velocity is based on the trajectory of the points of bitangency (POB), that is, the top and bottom most points of the ellipsoids (with respect to the vertical direction). I started with a certain angular velocity for the ellipse, typically 4°/view; for each ellipsoid I computed the angular velocity which would make the trajectories of its points of bitangency equal to those of the ellipse. The rationale behind this algorithm is the hypothesis that the human visual system uses the points of bitangency to compute the object's motion. Evidently, this is not the case because ellipsoids appear to slow down as they get thicker even though they are rotating at the same angular velocity. An alternative hypothesis is that the perceived angular velocity is proportional to the length of the POB trajectories obtained during a fixed time period. Equating these lengths results in objects that appear to be rotating at the same angular velocity. I will return to the issue of perceived angular velocity in the discussion section and in Experiment 6.2.

This procedure for computing angular velocity also equates image velocities for the different ellipsoids, where image velocity is defined as horizontal displacement at a point on the contour when going from one view to the next. Column **Max V** in Tables B.1 through B.8 shows the maximal velocity attained at any time during the sequence, and column **Mean V** shows the mean velocity averaged over all views and all points on the contour. Both maximal and mean velocity decrease somewhat with the ellipsoids' thickness. As the standard deviation in the image velocity is on the order of 60% of the mean, this decrease does not appear to be very significant. Also note that the standard deviation is about 60% for all ellipsoids, regardless of their thickness.

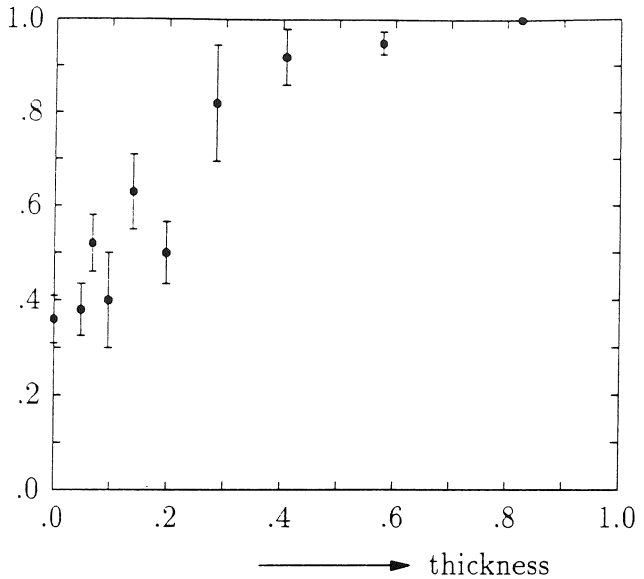
The fact that mean image velocity varies little with ellipsoidal thickness is probably the reason ellipsoids are perceived as rotating at the same velocity (section 6.2). In any event, this angular velocity adjustment eliminates the 2D cues of mean and maximal image velocity, and of the POB trajectories.

Apparatus. Stimuli were displayed on an IMI-455N. Subjects viewed the display in a dark room through a viewing tube which limited the field of view to approximately 10° of visual angle. The distance between the subject's eye and the screen was .9m.

Procedure. Subjects were instructed to indicate whether they perceived a flat object or one that was volume-containing, that is, not flat, when looking at a stimulus. Viewing time was unlimited. No feedback was given. A few seconds after the subject responded, a new stimulus appeared on the screen. The order of presentation was randomized so that each trial was equally likely to be of a flat ellipse or of any of the fat ellipsoids. Sessions lasted about 30 minutes.

Results and discussion:

Appendix B shows the details and outcome of a number of experiments whose main features are summarized in Table 6.1. The main result is that ellipsoids with thickness > .3 can be reliably distinguished from ellipses (Figure 6.7, and column **Thrhold thick** in Table 6.1). This is due to the fact that fat ellipsoids are perceived consistently as solid, "volume



thickness	d'
0.048	-0.27 ± .39
0.068	0.41 ± .34*
0.097	0.12 ± .37
0.139	0.71 ± .35*
0.198	0.37 ± .34*
0.283	1.27 ± .39*
0.405	1.75 ± .47*
0.578	2.02 ± .67*
0.826	5.57 ± .12*

Figure 6.7. Mean probability of responding "solid object" as a function of the ellipsoid's thickness in experiments B-1 through B-6 combined, and mean d'-s. Error bars indicate standard deviation; * indicates significance at 5%.

Table 6.1 Main parameters of experiments B-1 through B-10. Max Thick indicates the maximal thickness, Nr Thick indicates the number of different ellipsoids used in an experiment. Major Axis is $2\sqrt{1/a_{33}}$. Object slant is the angle between an ellipsoid's major axis and the vertical. Axis Slant is the angle between the axis of rotation and the vertical. Thrhld Thick is the minimal thickness which makes an ellipsoid discriminable from a flat ellipse.

Exp	Render	Nr Views	Max Thick	Nr Thick	Major Axis	Object Slant	Axis Slant	Thrhld Thick
B-1	contour	15	.826	10	4.9	60°	0°	.139
B-2	silhtte	15	.578	9	4.9	60°	0°	.139
B-3	contour	15	.578	9	4.0	60°	0°	.139
B-4	contour	15	.405	8	4.9	60°	0°	.283
B-5	contour	15	.405	8	4.9	60°	0°	.405
B-6	contour	15	.578	7	4.9	60°	0°	.578
B-7	contour	9	.578	9	4.9	60°	0°	.283
B-8	contour	15	.578	9	4.9	45°	0°	.283
B-9	contour	15	.826	8	4.9	60°	-22.5°	.283
B-10	contour	15	.826	8	4.9	60°	22.5°	.198

containing" objects. Phenomenally, these ellipsoids look solid; the percept also appears to be more stable than for the flatter ellipsoids and the ellipse. The latter elicit a much more ambiguous percept: Ellipses are sometimes perceived as solids (36% of the ellipses in experiments B-1 through B-8 are perceived as solids; the fat ellipsoids are almost never seen as flat).

The fact that flat ellipses are often perceived as solid makes it hard to fit a psychometric curve of the form $p(x) = 1 - \gamma exp(-(x/\alpha)^{\beta})$ to the data. As the experiment was of the yes/no type, $\gamma = 1.$, which means that $p(0) = 0$. This is clearly not the case as the flat ellipse is perceived as a solid object with a probability of .36 in experiments B-1 through B-8. One possible remedy is to choose some noise level, say 5%, at which to compute the discrimination measures. This ensures that we will never encounter $p(0)$ as all discrimination measures are now nonzero for the ellipse (and of course all ellipsoids). Figure 6.8 shows the result of fitting $p(x)$ in a least-squares sense with the motion discrimination measure as independent variable; the psychometric functions for the nonrigidity and rim prediction measures are very similar and are listed in Figure B.1 of Appendix B. Table 6.2 lists the thresholds (α), slopes (β), and errors for the motion, nonrigidity, and rim prediction discrimination measures. Each was computed using either the actual angular velocity of the rotating ellipsoids, that is, the angular velocity used to generate the displays, or the angular velocity estimated from the displays (section 6.2). In addition, the 2D cue of contraction has been treated as if it were a discrimination measure (angular velocity is of course not an issue now).

In addition to fitting psychometric curves, I also used a signal detection paradigm (Falmagne, 1985) to analyze the results. The trials containing solid ellipsoids served as signal trials (note that each level of ellipsoidal thickness is considered separate; i.e., the ellipsoids are not grouped together into one signal group; the number of different thicknesses is called NrLevels). The trials containing the flat ellipses served as noise trials. The hit rate is the fraction of "solid object" responses in the signal trials; the false alarm rate is the fraction of "solid object" responses in the noise trials. In each experiment, the number of signal trials at a particular thickness is 10; the number of noise trials is 10 $\times$ NrLevels (the number of ellipses equals the number of ellipsoids in an experiment). Thus, Each d' is based on 10 signal trials and 10 $\times$ NrLevels noise trials. The significance of the d' scores was determined using Marascuilo's (1970) one-signal significance test. The d' scores and their significance are shown in column d' in the Tables B.1 through B.10; columns Flat and $\neg$Flat list the actual number of "flat" and "solid" responses. Table 6.1 also lists the minimal ellipsoidal thickness with a significant d' score.

I draw the following conclusions from the results of these experiments:

(i) All experiments show the same basic result: the probability of perceiving a solid object increases with the ellipsoid's thickness. Thresholds vary from experiment to experiment, but ellipsoids with thickness $T > .4$ are consistently perceived as solid.

(ii) Silhouettes yield results comparable to those of contours (exp. B-2 vs. B-1). Thresholds for the psychometric functions are similar, but the slopes for silhouettes are about twice those for contours, suggesting that silhouettes yield more discriminable percepts. A possible explanation is that, in silhouette sequences, there is less of a tendency to assume correspondence across views and hence to impose a planar percept (this would imply that the perceived angular velocity of silhouette sequences is higher than of contour sequences; see

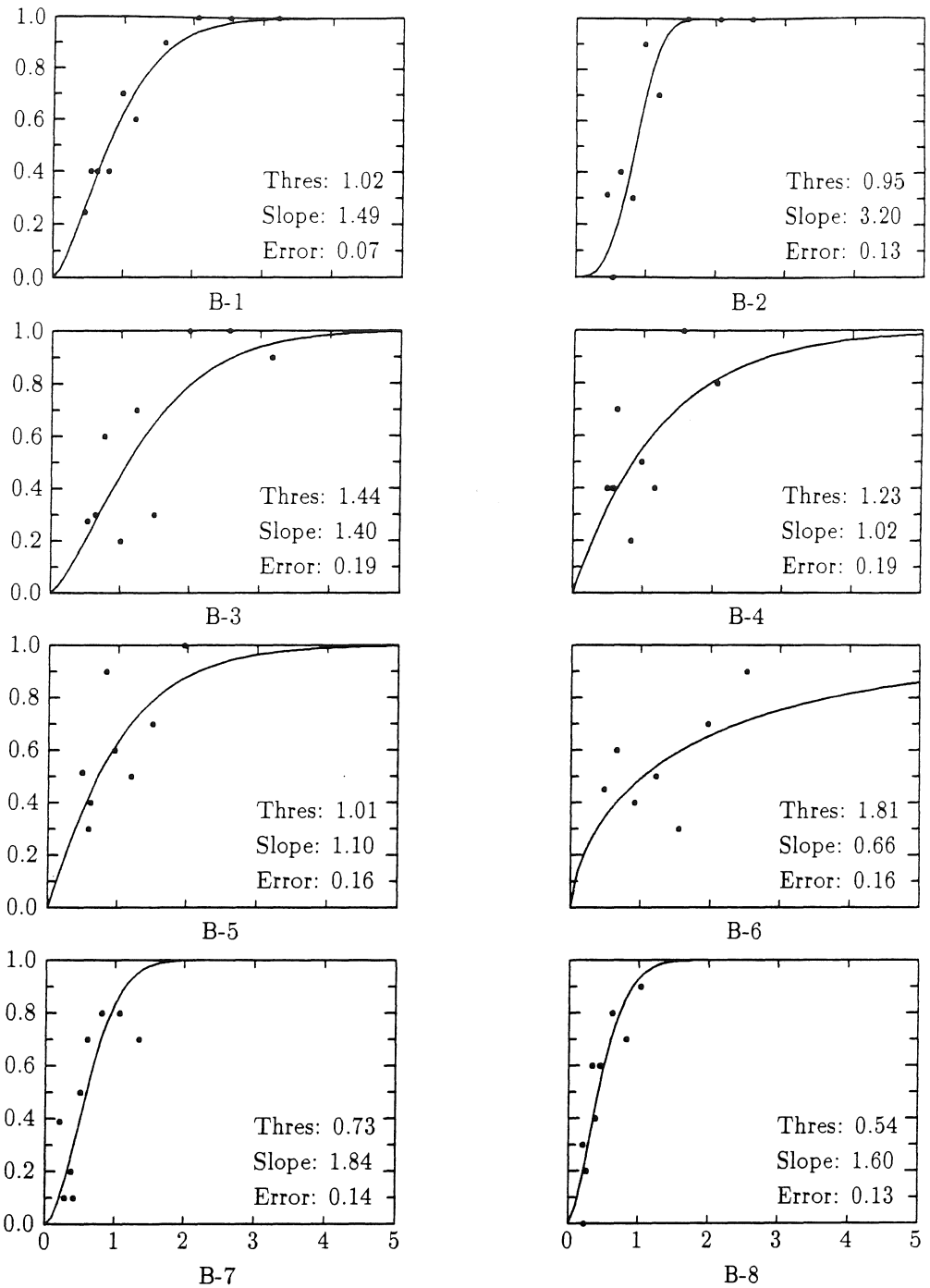


Figure 6.8 Psychometric functions with the *motion* discrimination measure as the independent variable. Used the *actual* angular velocity. Measures were computed using 5% noise. Thresholds are absolute thresholds. (See column Motion in Tables B-1 through B-8.)

Table 6.2 Thresholds and slopes for the motion, nonrigidity, and rim and tangent prediction discrimination measures, as well as for contraction. Abs. Thr. is the threshold derived from the absolute value of the discrimination measures. Rel. Thr. is the threshold derived from the relative increase of the discrimination measures with respect to those of the ellipse. Error is $(\frac{1}{n}\sum(p_e(x) - p_d(x))^2)^{1/2}$, where $p_e(x)$ is the estimated probability and $p_d(x)$ the actual probability of responding "not flat" to a stimulus, and n is the number of different ellipsoids in the experiment. In computing the discrimination measures, 5% noise was used.

	Experiments							
	B-1	**B-2**	**B-3**	**B-4**	**B-5**	**B-6**	**B-7**	**B-8**
Motion								
Actual θ								
Abs. Thr.	1.02	.95	1.44	1.23	1.01	1.81	.73	.54
Rel. Thr.	2.25	2.09	2.74	2.51	2.06	3.73	3.32	2.57
Slope	1.49	3.20	1.40	1.02	1.10	.66	1.84	1.60
Error	.07	.13	.19	.19	.16	.16	.14	.13
Estimated θ								
Abs. Thr.	.76	.68	.99	.94	.73	1.42	.74	.35
Rel. Thr.	1.59	1.43	1.89	1.84	1.43	3.20	3.53	1.68
Slope	1.49	4.28	1.53	1.02	1.35	.70	.92	2.3
Error	.06	.13	.18	.19	.15	.14	.19	.13
Nonrigidity								
Actual θ								
Abs. Thr.	1.98	1.85	2.29	2.22	1.92	3.02	1.73	1.13
Rel. Thr.	1.75	1.63	1.88	1.96	1.70	2.63	1.54	1.50
Slope	2.22	5.45	2.45	1.66	1.23	.81	2.92	2.73
Error	.03	.10	.15	.16	.18	.16	.15	.15
Estimated θ								
Abs. Thr.	1.53	1.43	1.65	1.52	1.44	2.14	1.70	.90
Rel. Thr.	1.35	1.34	1.38	1.37	1.27	1.88	1.50	1.13
Slope	3.0	9.7	3.5	3.1	2.1	1.2	2.1	5.20
Error	.03	.08	.15	.16	.17	.13	.20	.14
Rim pred.								
Actual θ								
Abs. Thr.	1.49	1.36	2.21	1.82	1.47	2.74	.86	80
Rel. Thr.	1.96	1.78	2.44	2.39	1.93	3.87	2.90	1.94
Slope	1.34	2.95	1.26	.88	.86	.59	1.46	1.49
Error	.06	.12	.19	.19	.16	.16	.13	.14
Estimated θ								
Abs. Thr.	1.05	.95	1.4	1.2	1.0	1.60	.55	.45
Rel. Thr.	1.35	1.23	1.55	1.56	1.27	2.26	1.74	1.15
Slope	1.65	4.22	1.62	1.06	1.32	.90	1.54	4.06
Error	.06	.13	.19	.19	.16	.14	.18	.13

Table 6.2 cont. Thresholds and slopes for the tangent prediction discrimination measure as well as for contraction.

	Experiments							
	B-1	B-2	B-3	B-4	B-5	B-6	B-7	B-8
Tangent pred.								
Actual θ								
Abs. Thr.	1.29	1.27	1.85	1.53	1.30	1.75	.72	.65
Rel. Thr.	1.33	1.33	1.46	1.53	1.30	1.81	1.55	1.39
Slope	1.75	3.26	1.46	.98	1.73	1.44	2.02	2.65
Error	.11	.20	.23	.21	.13	.14	.16	.09
Estimated θ								
Abs. Thr.	1.45	1.36	2.13	1.73	1.39	2.11	.96	.64
Rel. Thr.	1.50	1.41	1.68	1.73	1.39	2.18	2.06	1.38
Slope	1.35	2.59	1.34	.89	1.39	1.03	1.12	2.54
Error	.08	.17	.20	.20	.13	.12	.18	.10
Contraction								
Abs. Thr.[1]	.069	.073	0.073	.040	.053	.023	.011	.073
Abs. Thr.[2]	1.31	1.27	1.27	1.60	1.47	1.77	1.89	1.23
Rel. Thr.[1]	1.66	1.49	1.49	2.61	2.14	4.53	2.79	1.42
Slope[2]	4.45	14.89	6.72	3.87	3.00	1.57	13.00	5.20
Error[2]	.04	.13	.11	.17	.19	.17	.16	.19

[1] Values correspond to actual contractions as listed in Tables B.1 through B.8.
[2] Values correspond to psychometric curves; in order to fit psychometric curves, original contraction values were converted into $100 \times (.2 - \text{contraction})$.

next section). d'-s are similar; d' in both cases becomes significant if thickness exceeds .139.

(iii) The discrimination measures based on Gaussian and mean curvature cannot account for the data because they do not increase monotonically with ellipsoidal thickness for the particular families of ellipsoids used in these experiments. The reason they do not is because $1/a_{11}$, the ellipsoids' axis in the horizontal direction, was varied as a function of thickness. Of all the discrimination measures, only Gaussian and mean curvature are influenced considerably by changes in $1/a_{11}$. Experiment B-3 (Table B.3) provides a good example. $\mathcal{C}_G$ is largest at thicknesses .1, .2, and .6, precisely where $1/a_{11}$ is maximal. However, only the ellipsoids with T=.6 are perceived as solid; moreover, ellipsoids with thicknesses .3 or .4 are perceived as solid even though $\mathcal{C}_G$ is much lower here ($\mathcal{C}_G(.3) \approx \mathcal{C}_G(.1)/2$). The contour prediction measure can also be discarded because it varies very little with thickness (see, for example, column Contour in Table B.1).

(iv) The discrimination measures based on curvature in the direction of motion, nonrigidity, and the rim and tangent plane prediction yield very similar results; their psychometric functions fit the data equally well (Table 6.2).

(v) Changing the slant of the (major axis of the) ellipsoids from 60° to 45° (exp. B-1 vs. B-8) causes the absolute thresholds for the motion, nonrigidity and the rim and tangent

prediction measures to be halved; the threshold for contraction is not affected. Relative thresholds (absolute threshold at thickness T divided by the measure for the ellipse, i.e., T = 0) for all five measures remain the same. The minimal thickness with significant d' is .283 for both slants.

It should be pointed out that this change in slant is not obvious perceptually. In fact, the ellipsoids, and especially the flatter ones including the ellipse, appear to be slanted very little. The perceived slant is more like 30° rather than the 45° and 60° used in generating the stimuli. See Chapter 5 for the perception of slant of rotating planar patches rendered by dots; slant is underestimated for these stimuli as well.

(vi) Decreasing the major axis of the ellipsoid from 4.9 to 4.0 (exp. B-1 vs. B-3) did not affect any of the psychometric curves appreciably; thresholds remain more or less constant (Table 6.2). Minimal discriminable thickness remains .139.

(vii) Decreasing the number of views (i.e., the range of views of the rotating ellipsoids) decreases all discrimination measures (exp. B-1 vs. B-7). The minimally discriminable thickness increases from .193 to .283. This makes sense because an observer can make finer distinctions if he has more information, in this case in the form of more views. According to Table 6.2, the absolute threshold for motion and nonrigidity remains the same, but drops significantly for the rim and tangent prediction measures and even more for the contraction measure (.067 to .011). Relative thresholds increase except for nonrigidity. The slopes of the psychometric curves do not show a clear pattern.

(viii) Contraction is a serious 2D confound; the psychometric curves based on it are reasonable (Figure B.1i), with only a slightly higher error than those of the 3D discrimination measures. However, if contraction were the sole determinant, or the most important one, then we would not expect to find a difference in experiments B-1 and B-2 which only differ in their rendering of the stimulus (contour vs. silhouette, resp.). As I mentioned under (ii), the difference might be attributed to a decreased tendency to impose correspondence in silhouette as compared to contour sequences.

If contraction were the main cue, one would expect performance to be much better in experiment B-6 in which contraction ranges from .105 down to .010, that is, the same range as in experiments B-1, B-2, and B-4. However, d'-s are much lower in B-6 than in the other three experiments, which is mainly due to an increase in the false alarm rate from 31% to 45%. Similarly, one would expect performance to be worse than it actually is in experiment B-8, as contraction now ranges only from .109 to .059. The only results that could be explained through contraction are B-4 and B-5: now performance decreases as the range of contraction shrinks from (.105,.010) to (.105,.032).

Summarizing, even though contraction covaries with response, there are a number of exceptions which indicate that response is unlikely to be solely based on contraction.

(ix) Slanting the axis of rotation by ±22.5° (exp. B-9 and B-10 vs. B-6) results in a decrease of the minimally discriminable thickness (.198 and .283 vs. .578). This effect is mainly due to the fact that the flat ellipses are more consistently perceived as flat. The false alarm rate drops from 45% to 19%, whereas the overall hit rate drops slightly from 57% to 49% if we group all ellipsoids together into one signal group. If we divide the ellipsoids into a moderately fat group (thickness $\leq$ .193) and a fat group group (thickness $>$ .193), then the hit rate for the moderately fat group drops from 50% to 28%, and the hit rate for the

fat group increases from 63% to 70%.

The effect of axis slant was further investigated in Experiment B-15. The family of ellipsoids was $\mathcal{E}_T : x^2/2 + y^2/T + z^2/6 = 1$, slanted $-60°$; there were 4 levels of thickness, namely, 0.0, .33, .67, and 1.0. The stimuli were rendered as silhouettes (dark figure, bright background). Each sequence consisted of 15 views separated by 4° of rotation; each view is shown for 100ms, and is then immediately replaced by the next one. Ellipsoids were scaled such that the vertical size of their images were the same at the moment of zero rotation (all contours in Figure 6.9 whose major axis is vertical have the same vertical extent). In each experimental session, the axis of rotation, which goes through the center of the ellipsoid, was either slanted towards the observer (top of axis towards the viewer; slant angles were $0°, 22.5°$, and $45°$), or away from the observer (slant angles $0°, -22.5°$, and $-45°$). Recall that the ellipsoids themselves were slanted by $-60°$; thus, an ellipsoid rotating about an axis slanted by $-60°$ is rotating about its major axis. The sequence of contours of an ellipsoid with thickness T rotating about an axis slant by ω will be referred to as $\mathcal{S}^T_\omega$. In each experimental session, each ellipsoid with nonzero thickness was shown six times for each slant condition; as before, the number of ellipsoids with nonzero thickness equaled the number of flat ellipses. Thus, there were 6x3x3x2= 108 judgments per session; each session lasted about 30 minutes. The stimuli are illustrated in Figure 6.9; for clarity only three views in one direction are shown. If the axis of rotation is not slanted, the vertical size of the ellipsoid's projection does not change; its POBs, which are fixed points, lie on a straight line. The vertical size of the projection increases (decreases) if the axis of rotation is slanted by $-22.5°$ or $-45°$ (22.5 or 45°); see also Figure 4.7, which shows the envelope of ellipsoids rotating about axes slanted by various degrees.

The results of these experiments are summarized in Table 6.3. In none of the experiments did the subjects have a clear impression of the orientation of the axis of rotation; there was a tendency to perceive the axis as frontoparallel. It is not known whether this is peculiar to dynamic occluding contours or whether it is generally true for structure-from-motion displays.

If the axis of rotation was slanted by 45°, neither subject could discriminate between ellipsoids and ellipses; all objects now appeared to be solid. If the axis of rotation was slanted by $-22.5°$ or $-45°$, JB could still discriminate between ellipsoids and ellipses.

A comparison of the contours under the various slant conditions can explain these results (Figure 6.9). Slanting the axis of rotation towards the observer ($-22.5°$ or $-45.0°$) changes the vertical size of the contours less than slanting it away. And what change there is occurs predominantly at the extremal views. The flat ellipses were still clearly perceived as flat. If, on the other hand, the axis of rotation is slanted away there is a considerable and gradual decrease in vertical size. Note that $\mathcal{S}^0_{45}$ is very similar to $\mathcal{S}^1_{45}$ and to $\mathcal{S}^1_0$ modulo a vertical size change. As a vertical size change could be interpreted as a change in viewing distance (under perspective projection), it is not surprising that $\mathcal{S}^0_{45}$ was perceived as solid.

These results confirm the hypothesis, developed in Chapter 4, that the slant of the axis of rotation cannot be recovered from dynamic occluding contours. Using points of correspondence it can only be determined whether the axis of rotation is frontoparallel or not.

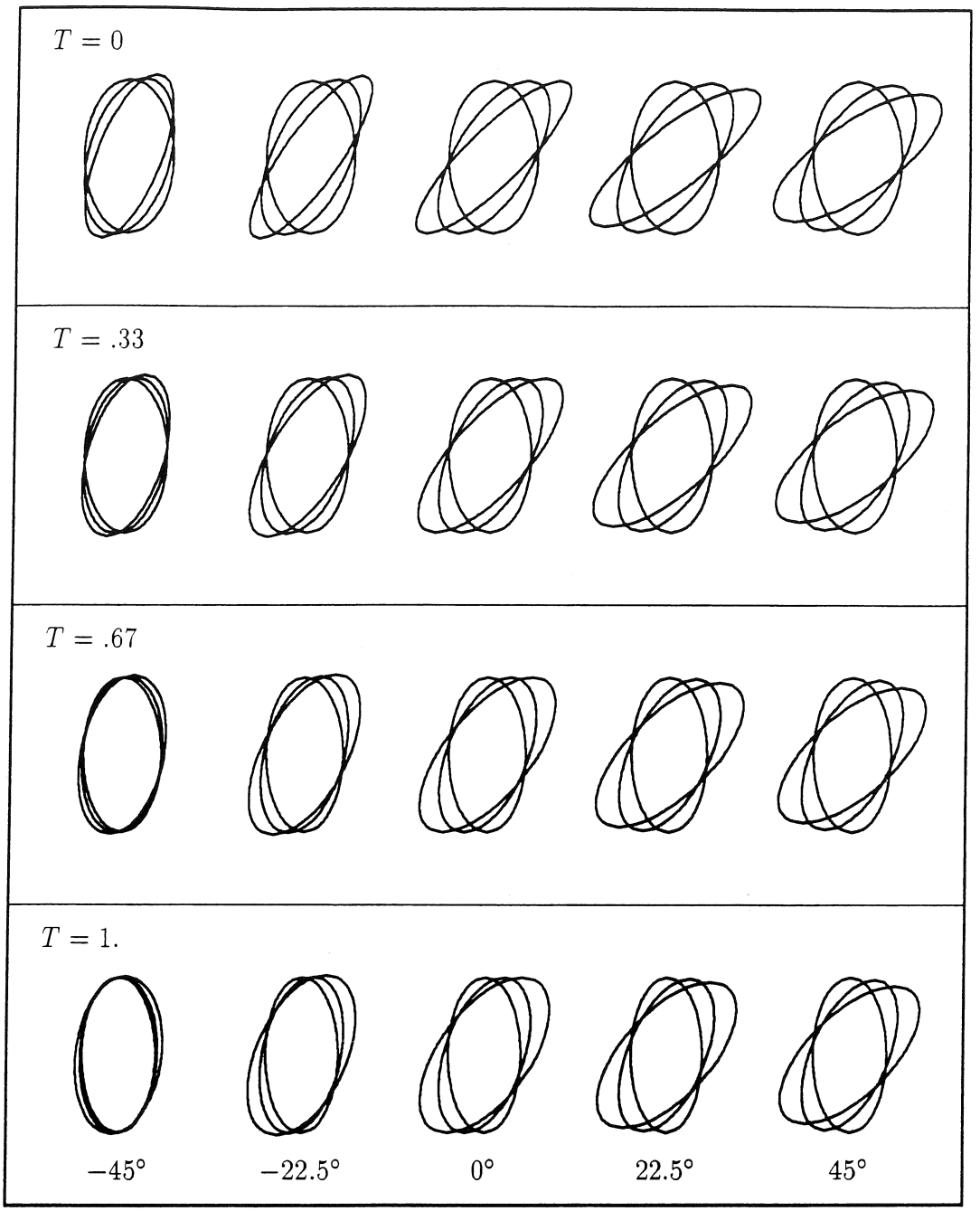


Figure 6.9. Three views each of four members of the family of ellipsoids $x^2/2+y^2/T+z^2/6 = 1$, whose major axis initially makes an angle of 60° with the vertical. Ellipsoids are rotated about 5 different axes of rotation; slant of axes is indicated by the bottom row of numbers. The top of axes slanted by -22.5° and -45° is away from the viewer, as is the top most part of the ellipsoids. The amount of rotation between views is 15°.

Table 6.3 Discrimination of solid ellipsoids and flat ellipses as a function of the slant of the axis of rotation. Numbers indicate the percentage of "flat" responses. The column labeled .33 – 1.0 refers to the average of "flat" responses for ellipsoids with nonzero thickness.

Subject	Axis Slant	Thickness .33	.67	1.0	.33–1.0	0.0
JB	0.0°	100	0	0	33	100
	22.5°	16	0	0	6	83
	45.0°	0	16	0	6	0
JT	0.0°	67	16	67	50	100
	22.5°	16	0	0	6	11
	45.0°	0	0	0	0	0
JB	0.0°	0	0	0	0	56
	−22.5°	50	0	0	17	100
	−45.0°	16	0	0	6	89
JT	0.0°	0	0	0	0	89
	−22.5°	50	17	0	22	100
	−45.0°	83	17	33	50	100

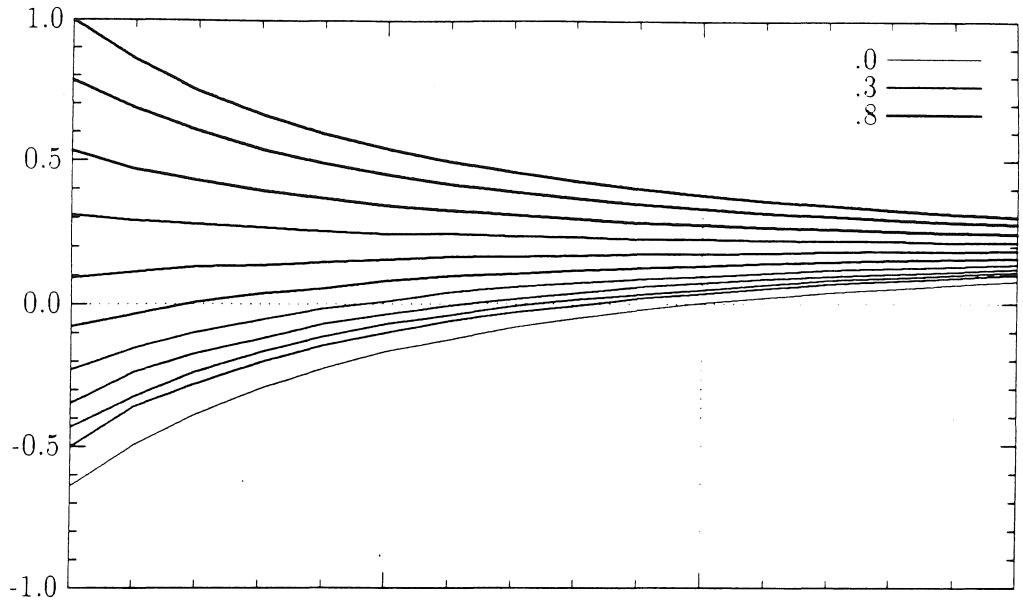
Returning to experiments B-1 through B-8,

(x) there is a considerable RMS error in all psychometric functions, especially for the moderately fat ellipsoids. This might be due to the fact that these ellipsoids can be perceived as solid or as planar depending on the perceived angular velocity. There is no such leeway for the fat ellipsoids; they can only be seen as solid. In the next section, I will explore the role of angular velocity in detail.

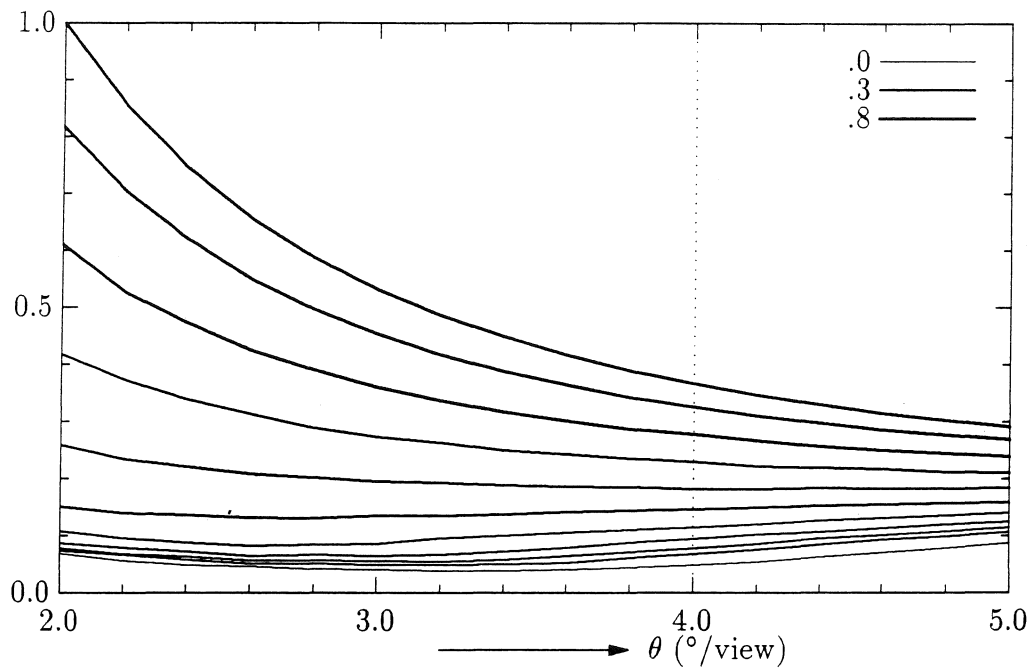
6.2 Perception of angular velocity of rotating ellipsoids

In the preceding section I mentioned that the thicker an ellipsoid the slower it appears to be rotating, even though the simulated angular velocity is kept constant. I will now further investigate the perception of angular velocity.

So far I have assumed that the human visual system somehow computes the correct angular velocity, which in principle is possible using the trajectories of fixed points. I used the actual angular velocity to compute the various discrimination measures in the previous

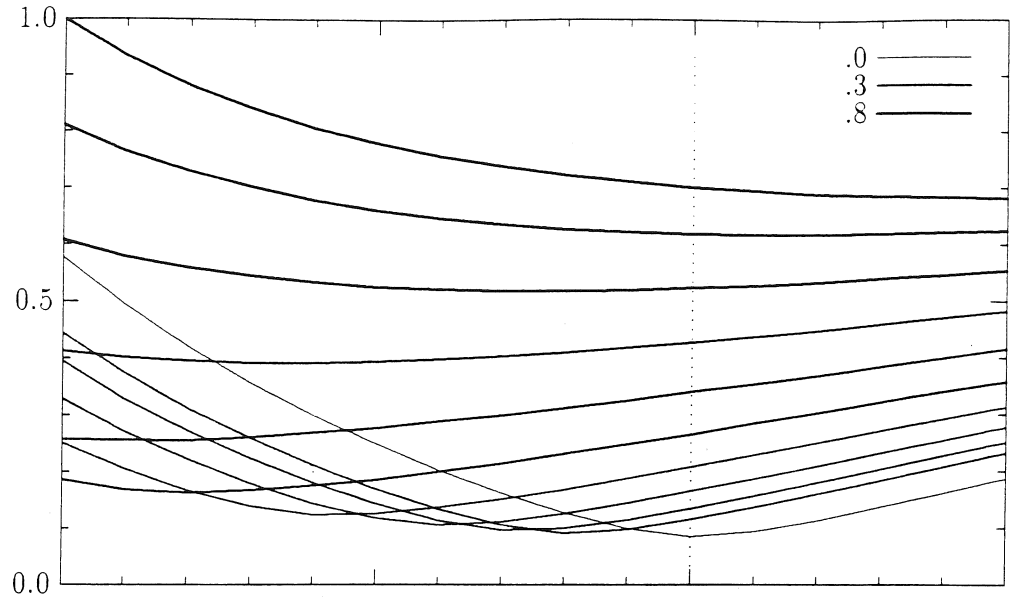


(*a*) Curvature in direction of motion (- and +)

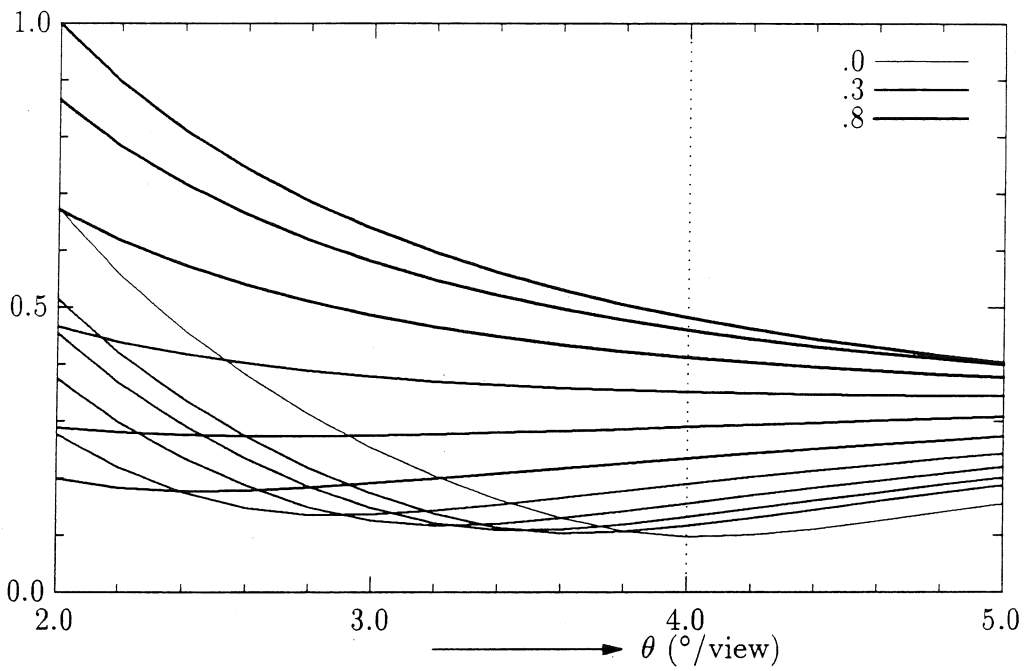


(*b*) Curvature in direction of motion (+ only)

Figure 6.10. Discrimination measures as a function of angular velocity θ. Stimuli were generated with $\theta = 4°$/view, and consisted of 15 views. Family of ellipsoids is as in Figure 6.2. Thickness of curves is proportional to the thickness of the ellipsoid. Ordinate indicates fraction of maximal value; values are means of 10 separate simulations at 5% noise; standard deviation is less than 10% of the mean.



(*c*) Rim prediction



(*d*) Nonrigidity

Figure 6.10 cont. (a) Curvature in direction of motion; curvature was allowed to assume negative values. (b) Curvature in direction of motion; negative curvature values were set to zero. (c) Rim prediction; predicted and actual rims are compared during 20°. (d) Nonrigidity measure.

experiment. What would happen to these measures if the perceived angular velocity (θ_{per}) were to deviate from the actual one (θ_{act})? Figure 6.10 shows how the motion, rim prediction, and the nonrigidity measures vary as a function of perceived angular velocity for the family of ellipsoids $\mathcal{E}_T : x^2+y^2/T+z^2/6 = 1$, slanted by 60° and rotating about the vertical at 4°/view. I used 5% noise in computing the measures (section 6.1). It is perhaps easiest to start with the nonrigidity measure (Figure 6.10d). Each curve in Figure 6.10d corresponds to a particular member of $\mathcal{E}_T$, and the curve's thickness increases with the member's thickness ($0 \leq T \leq .8$). Consider the planar ellipse, represented by the thinnest curve; it has a minimum when $\theta_{per} = \theta_{act}$. Thus, the most rigid 3-D interpretation is obtained if the actual angular velocity is used to compute relative depth via the relationship $z = -dx/d\theta$; underestimation and overestimations of angular velocity both lead to less rigid interpretations. As the thickness of the ellipsoid increases, the minimum shifts towards lower and lower perceived angular velocities and the value of the nonrigidity measure at the minimum increases. For $T > .25$, the minimum virtually disappears and the nonrigidity measure is large for all angular velocities.

The rim prediction measure behaves similarly (Figure 6.10c). It also has a clear minimal for $T < .25$, which occurs for $\theta_{per} < \theta_{act}$; it also does not have a clear minimum for $T > .25$.

The situation for the discrimination measure based on surface curvature in the direction of motion, κ_m, is slightly more complex. For opaque objects, $\kappa_m > 0$ (Koenderink, 1984). If we use the actual angular velocity to compute κ_m using the equation $1/\kappa_m = x + \partial^2 x/\partial\theta^2$, then $\kappa_m > 0$, but if we use the wrong angular velocity, κ_m can and does become negative. The question is whether to allow κ_m to become negative or to clip it at zero. As there is no compelling a priori reason to prefer one over the other, I looked at both cases. Figure 6.10b shows what happens when negative values are simply ignored. There are no clear minima in the measure for any of the ellipsoids, even for the flat ellipse, whose rather shallow minimum actually occurs at 3.3°/view instead of at the actual 4°/view. Figure 6.10a shows what happens if negative values for κ_m are included. Of interest now are the zero-crossings of the measures, i.e., the angular velocity at which the measure is zero. The zero-crossing for the flat ellipse occurs at θ_{act}. As thickness increases (up to to about .3), the zero-crossing occurs at progressively lower θ_{per}, analogous to what happened with the nonrigidity and rim prediction measures.

Summarizing, the most rigid 3-D interpretation for moderately fat ellipsoids occurs at angular velocities that are less than the actual one (i.e., the one used to generate the sequence of occluding contours). In other words, the contours of a rotating, moderately fat ellipsoid can be interpreted as those of a flat ellipse which is rotating slower than the ellipsoid. Preferring such an interpretation can be viewed as preferring an interpretation that assumes correspondence across views. Indeed, applying the structure-from-motion algorithm of Equation 4.18 to the contours of a rotating ellipsoid yields angular velocities that decrease with the ellipsoid's thickness (Figure 6.11a; contours are treated as consisting of 120 points; all pairs of neighboring points are followed over time and used to compute angular velocity; the angular velocity graphed is the mean for all such pairs averaged over all views). Equation 4.18 also contains a sanity check on the computed angular velocity, namely that $|\cos\theta| \leq 1$. Figure 6.11b shows the average portion of the contours for which this check fails, again as a function of ellipsoidal thickness. As expected, failure rate increases with thickness, reaching

80% for the fattest ellipsoids. Note that a moderate amount of noise (5%) hardly affects the computed angular velocity or the failure rate. Computed angular velocity increases if $T > .27$, but that is highly suspect because of the accompanying high failure rate.

Figure 6.11a (dotted line) indicates the angular velocity at which the rim prediction measure achieves its minimum (the motion and nonrigidity measures behave virtually the same as the rim prediction measure). It follows the same course as the angular velocity computed assuming correspondence across views, although its slope is considerably steeper.

In the next experiment, I will investigate how the human visual system perceives the angular velocity of rotating ellipsoids from their contours.

Experiment 6.2: Estimating the angular velocity of ellipsoids rotating about a frontoparallel axis.

The purpose of this experiment is to determine how the thickness of a rotating ellipsoid influences its perceived angular velocity.

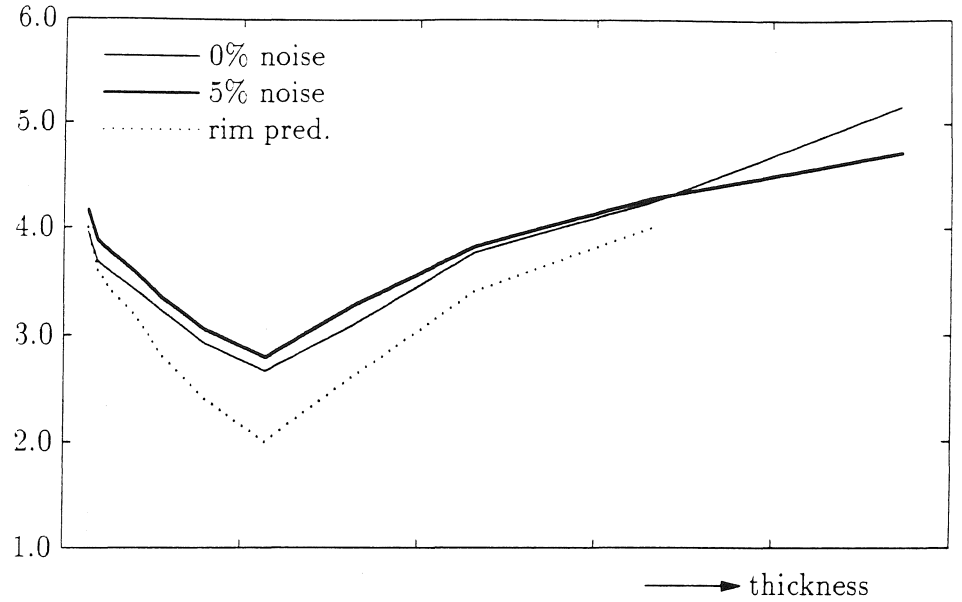
Methods:

Subjects. Three three subjects were the author (JB), JT, and JL. All three were familiar with psychophysical experiments. JT and JL were naive as to the purpose of this experiments. JL had considerable experience in dynamic probe dot experiments.

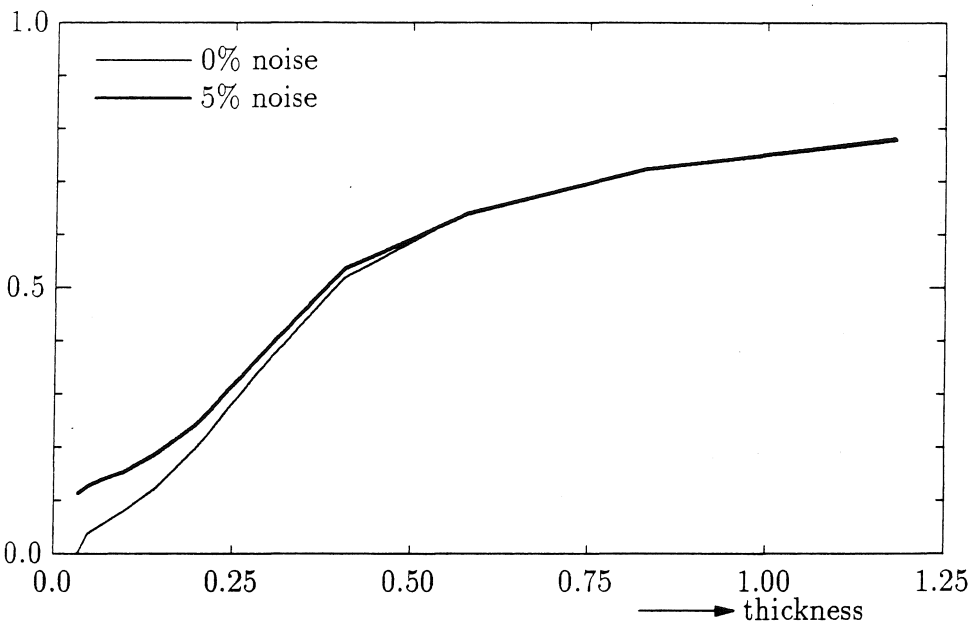
Design. There were two independent variables, namely the thickness of the ellipsoid (8 levels, from 0.0 to 1.0), and the simulated angular velocity (2 levels: 3°/view, 21 views, and 4°/view, 16 views). Each ellipsoid was shown 5 times rotating at a particular angular velocity. Thus, there were 8x2x5 = 80 angular velocity estimates in an experiment. The sequence of presentation was randomized.

Stimuli. Each visual display consisted of two components. On the left was shown a sequence of contours of an ellipsoid rotating about a vertical axis at constant angular velocity θ. Ellipsoids were chosen from the family $\mathcal{E} : x^2/2 + y^2/T + z^2/6 = 1$; major axis of ellipsoids made an angle of 60° with the vertical. Subjects saw a sequence $v_1 v_2 ... v_{n-1} v_n v_n v_{n-1} ... v_2 v_1 v_1 v_2 ...$, corresponding to an ellipsoid rotating back and forth. The range of rotation was $[-30°, 30°]$. Each view was shown for 100ms, and was then immediately replaced by the next one (OSA = 100ms; ISI is zero). The direction of rotation reversed instantaneously at v_1 and v_n in the sequence. Contours were computed using orthographic projection. *Contours* were rendered as bright curves on a dark background. *Silhouettes* were contours filled in by horizontal line segments giving the impression of a filled-in figure. *Control* stimuli consisted of about 100 dots distributed over that part of an ellipsoid's surface that faces the subject, yielding an opaque, convex surface; the implicit outlines of these stimuli were similar to those of the Contours and Silhouettes.

On the right was shown the so-called angular velocity probe, 50 dots randomly distributed on a simulated transparent cylinder with a vertical axis and rotating about the vertical. The vertical size of the probe was the same as that of the contours (10cm); its horizontal size was comparable to that of the ellipsoids. The same probe was used in all experiments. The angular velocity of the probe could be adjusted interactively by a joystick; by moving the joystick to the left (right), angular velocity decreased (increased). The initial angular velocity of the probe was chosen at random between 1°/view and 6°/view. The reason for choosing two different angular velocities rather than one was to decrease the possibility of subjects



(*a*) Angular velocity



(*b*) Error fraction

Figure 6.11. Estimated angular velocity as a function of the ellipsoids' thickness. The family of ellipsoids is as in Figure 6.2. (a) Lines indicates angular velocity (°/view) as estimated from a sequence of occluding contours assuming correspondence between successive views; values are means of 10 simulations at 5% noise (thick line) and without noise (thin line). The dotted line specifies the angular velocities at which the rim prediction measure shown in Figure 6.10c is minimal. (b) The average portion of the contours for which the angular velocity computation fails (thick line: 5% noise; thin line: no noise).

maintaining a mental copy of a particular setting of the probe which might introduce a bias in their responses.

Apparatus. Stimuli were displayed on an IMI-455N. Subjects viewed the display in a dark room through a viewing tube which limited the field of view to approximately 10° of visual angle. The distance between the subject's eye and the screen was .9m.

Procedure. Subjects were instructed to adjust the angular velocity of the probe until it matched that of the rotating object on the left. After the angular velocity adjustment, the subject indicated whether a flat or a fat object was perceived by moving the joystick to the left or right (no feedback as to the accuracy of the latter judgment was given). Subjects could take as long as they felt necessary. Each experimental session lasted about an hour.

Results and discussion:

Figures 6.12 and 6.13 show the results of the angular velocity judgments in the form of the mean and standard deviation of the responses and a straight line fitted in the least-squares sense. Figure 6.13 also indicates the fraction of objects that were perceived as flat (summed over the 3°/view and 4°/view responses). Table 6.4 lists the y-intercepts and slopes of the lines fitted to the velocity judgments. The y-intercept is actually called θ_0 because it is the perceived angular velocity of the planar ellipse having zero thickness; slopes have been normalized by dividing them by θ_0. The significance of predicting perceived angular velocity from thickness is quantified by p_F.

All four subjects agreed that the task was very hard, especially for the contours and silhouettes. Equating the angular velocity of two different objects rotating about the vertical is not as simple as equating say their linear velocity. It seems that each subject developed some private strategy and criterion for performing the task. This would account for the substantial differences even in the controls having a lot of easily identifiable, fixed points; for example, θ_0 for the ellipsoidal surface patch rotating at 3°/view was judged as 2.6, 5.4, 4.5, and 3.5°/view. A similar range was found for the patch rotating at 4°/view, namely, 2.9, 5.6, 4.5, and 3.9°/view; it is interesting that the actual 1°/view difference in angular velocity is not reflected in the judgments (average θ_0 for the 3°/view condition was 4.0 ± 1.1, and for 4°/view it was 4.2 ± 1.0). On the other hand, with only one significant exception (VB, 4°/view), the slopes for the controls were near zero indicating that the subjects were at least consistent in their strategies. All subjects reported tracking one of the two POBs of the contour or silhouette and comparing its motion with a point on the cylinder.

The data can be summarized as follows.

Perception of shape from contours and silhouettes. All subjects saw the objects slanted backwards (top of object farthest away). The perceived slant of the objects ranged from 10° to 60°, so that there is a considerable underestimation of the "actual" slant of 60°. Only subjects JB and JT were able to distinguish the flat ellipses from the fat ellipsoids; both clearly perceived solid objects with $T \neq 1$, the amount of volume/curvature increasing with thickness. VB and JL perceived most objects as flat and rigid, nonrigid otherwise; they never saw a solid object even after being told about what was being simulated.

Perception of angular velocity using contours. Perceived angular velocity decreased with thickness for all subjects. Averaged over the four subjects and the two angular velocities, slope is $-.30 \pm .08$. Averaged over JB and JT, the slope is $-.25 \pm .05$; averaged over JL and VB, the slope is $-.35 \pm .08$. Thus, the average slope of the two subjects who perceived

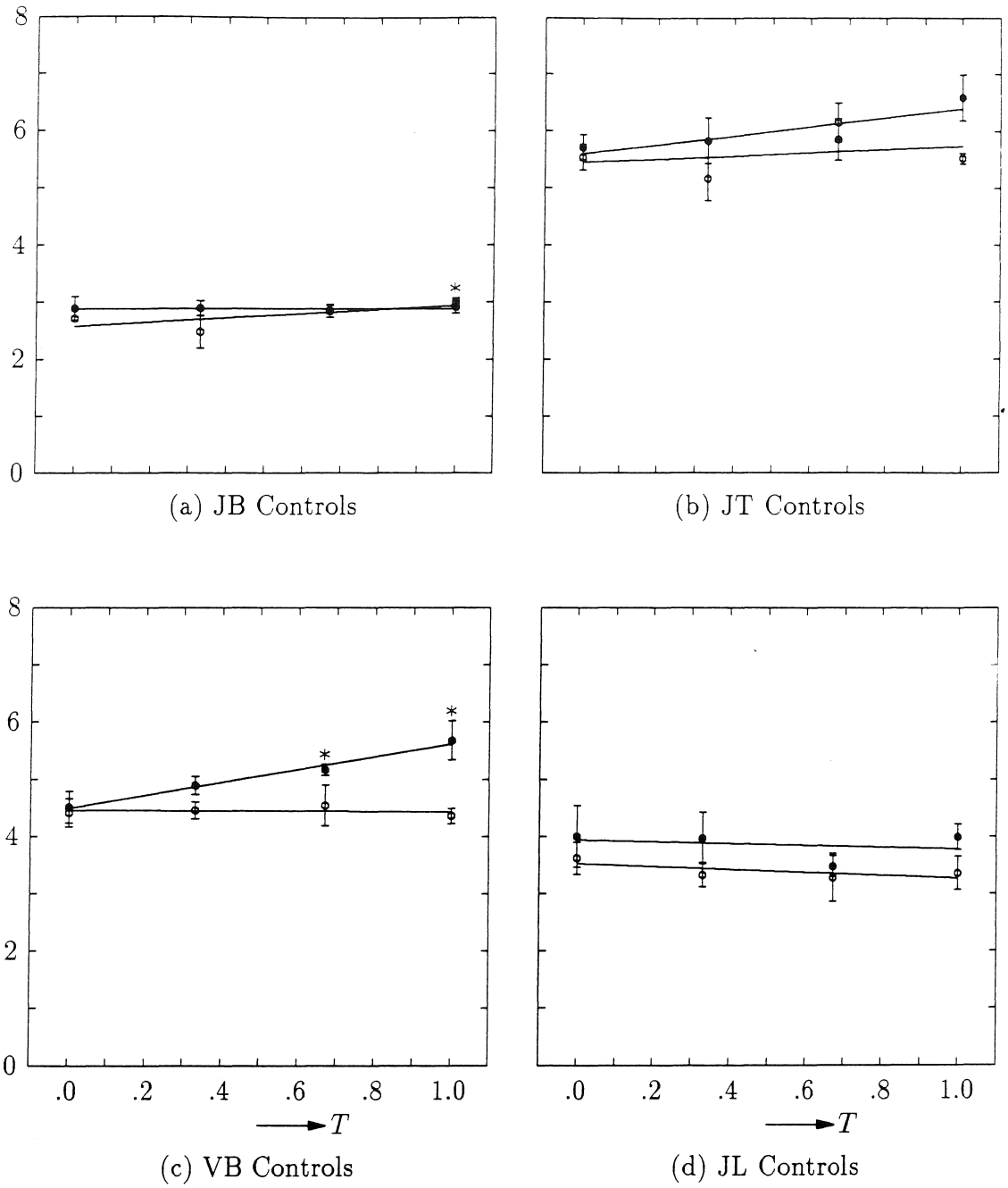
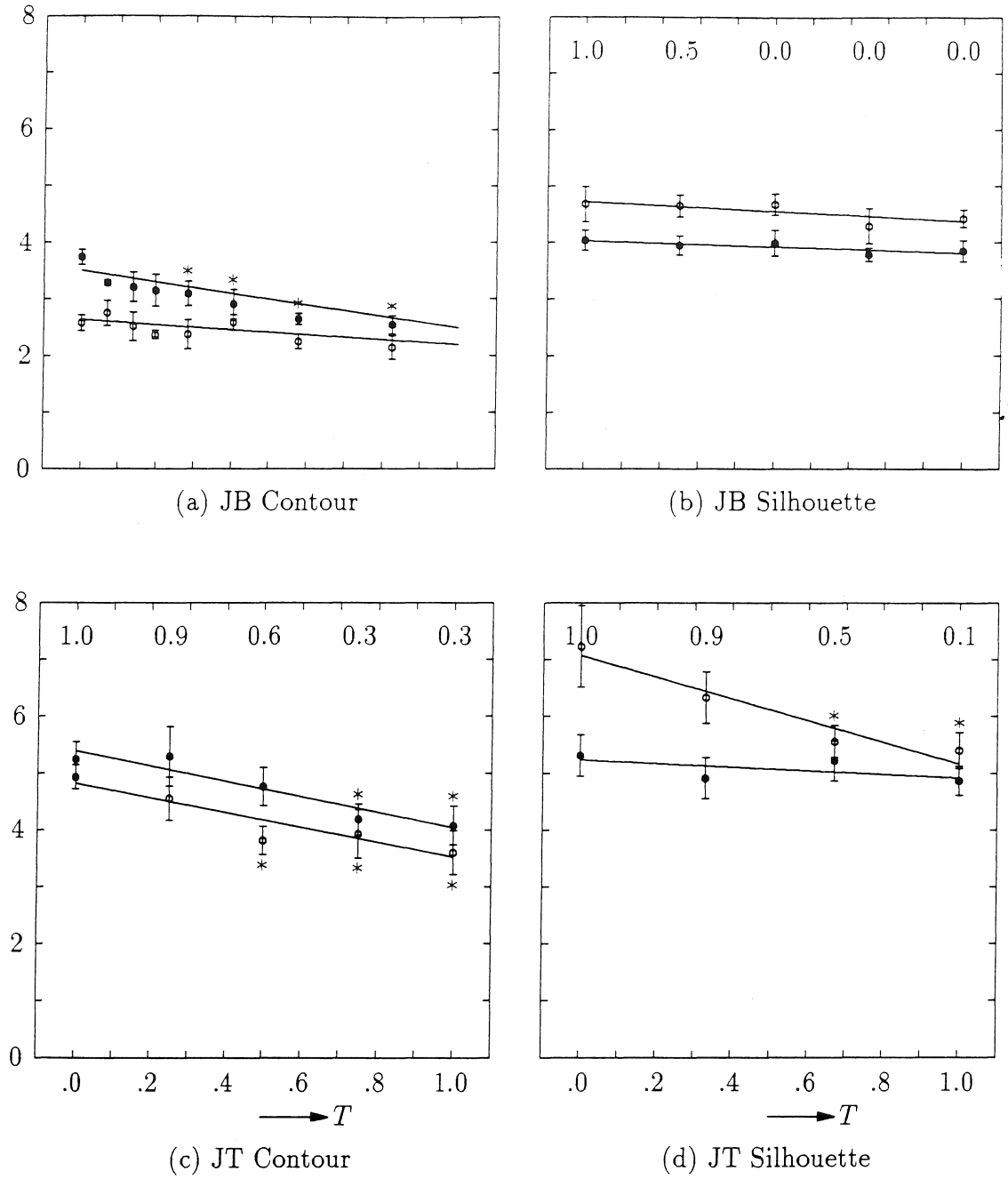


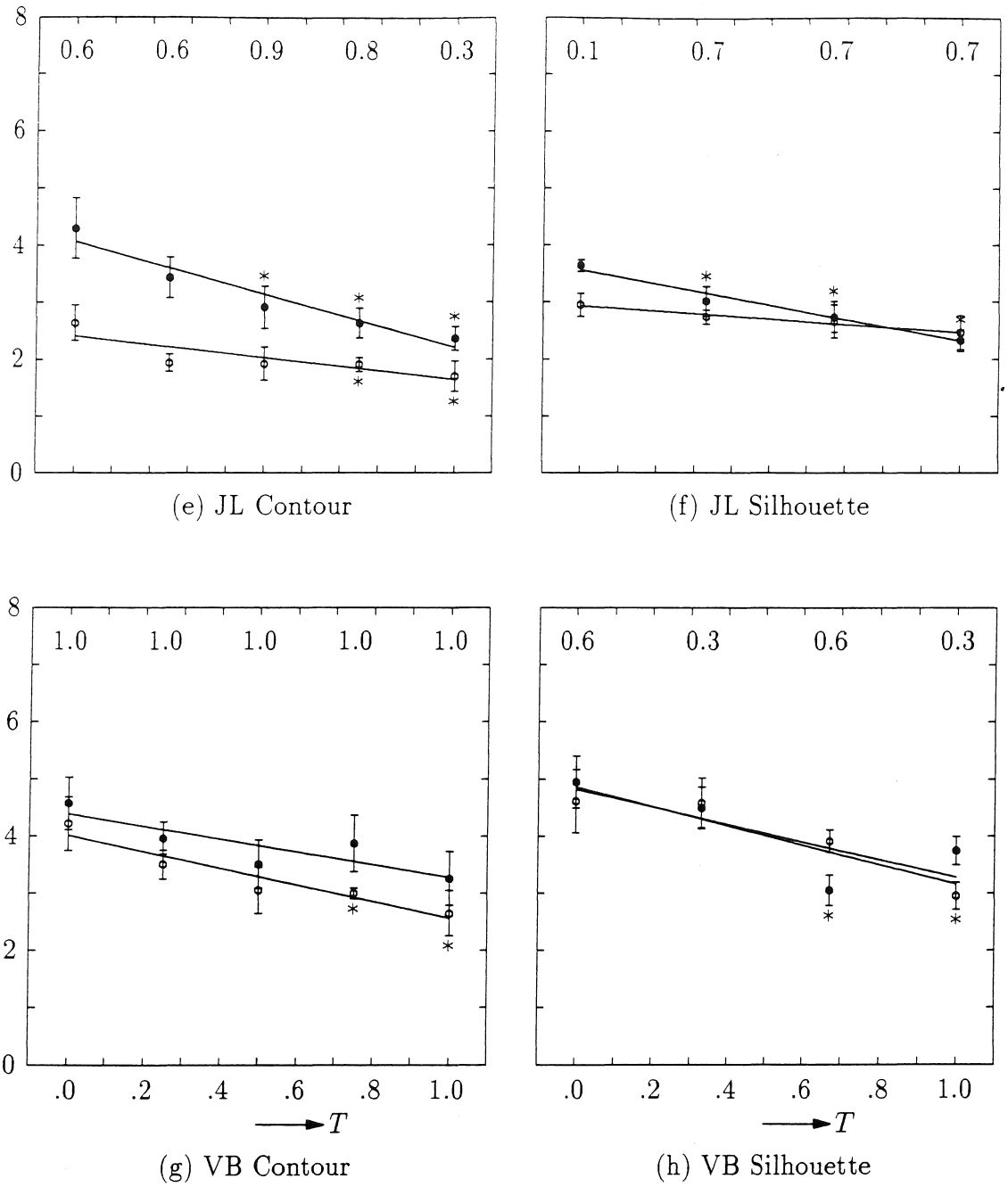
Figure 6.12. Angular velocity judgments of control stimuli for subjects JB, JT, JL, and VB. Error bars indicate standard deviation. * indicates that the angular velocity judgment at that thickness differs significantly from judgment at zero thickness (5% significance level). See Table 6.4 for a listing of a slopes and y-intercepts (θ_0).



(a) JB Contour (b) JB Silhouette

(c) JT Contour (d) JT Silhouette

Figure 6.13. Angular velocity judgments for subjects JB, JT, JL, and VB. Error bars indicate standard deviation. Numbers at top of graph indicate the fraction of objects perceived as flat at that thickness (10 trials each). * indicates that the angular velocity judgment at that thickness differs significantly from judgment at zero thickness (5% significance level). See Table 6.4 for a listing of a slopes and y-intercepts (θ_0).



(e) JL Contour (f) JL Silhouette

(g) VB Contour (h) VB Silhouette

Figure 6.13.cont.

Table 6.4 Angular velocity judgments. θ_0 is the judged angular velocity for an ellipsoid with zero thickness. $\theta_T/\theta_0 = (\triangle\theta/\triangle T)/\theta_0$, that is, the normalized slope of the line fitted in a least-squares sense to the angular velocity judgments expressed as a function of thickness T. p_F indicates significance of this prediction of angular velocity from thickness. Under 2D cues are listed the normalized slopes that would result if response had been proportional to mean image velocity, maximal image velocity, maximal extent of the projection, length of the POB trajectories, and the area of the projection summed over all views.

		3°			4°		
Display	**Subject**	θ_0	θ_T/θ_0	p_F	θ_0	θ_T/θ_0	p_F
Control	JB	2.58	.14	.30	2.88	.00	.90
	JT	5.44	.05	.79	5.59	.14	.14
	VB	4.45	.00	.90	4.49	.25	.01
	JL	3.52	-.07	.30	3.94	.04	.70
Contour	JB	2.64	-.17	.02	3.51	-.29	.00
	JT	4.82	-.27	.02	5.39	-.25	.01
	VB	4.01	-.36	.01	4.38	-.25	.06
	JL	2.41	-.32	.07	4.07	-.46	.01
Silhouette	JB	4.72	-.07	.12	4.03	-.06	.09
	JT	7.07	-.27	.04	5.23	-.06	.40
	VB	4.86	-.35	.06	4.82	-.32	.21
	JL	2.93	-.16	.02	3.56	-.35	.01
2D Cues	Mean Vel		-.44			-.44	
	Max Vel		-.43			-.43	
	Max Ext		-.17			-.17	
	POB Traj		-.31			-.31	
	Area		-.12			-.12	

solid objects was significantly less than the average slope of the subjects who saw only flat or nonrigid objects.

Perception of angular velocity using silhouettes. Perceived angular velocity remained constant for JB (3° and 4°/view) and JT (4°/view), and decreased with thickness for JL and VB. Averaged over the four subjects and the two angular velocities, slope is $-.21 \pm .12$. Averaged over JB and JT, the slope is $-.12 \pm .09$, a significant decrease with respect to the Contour condition. Averaged over JL and VB, the slope is $-.30 \pm .08$, which does not significantly differ from the Contour condition. A strange effect is that the slower objects result in angular velocity judgments that equal (JL and VB) or exceed (JB and JT) those for the faster objects. This is not due to any difference in perceived shape between the 3° and 4°/view conditions.

From these results I draw the following conclusions:

(i) Subjects JB and JT, who clearly perceived solid ellipsoids and who could reliably distinguish between the flat ellipses and solid ellipsoids, had lower slopes than subjects VB and JL, who did not perceive solid objects and who could not make the distinction. This is especially true for the silhouette condition in which slopes for JB and JT are zero, indicating that these subjects could perceive and respond to the actual angular velocity of the rotating ellipsoids. The fact that their slopes for contours are lower than for silhouettes could be explained by assuming that the visual system tends to assume correspondence more for contours than for silhouettes (the former can be viewed as wire frame figures) and computes angular velocity on that basis. As shown in Figure 6.11a, this has the desired effect of lowering θ_{est} with thickness (at least up to a thickness of about .25). The resulting normalized slope for the portion of Figure 6.11a in which angular velocity decreases is $-.80$, which is clearly much steeper than the observed ones. Other algorithms for computing angular velocity might lead to a closer match.

An alternative explanation for the difference between JB and JT on the one hand, and VB and JL, on the other, lies with 2D image cues. Perhaps VB and JL simply responded according to some 2D image characteristic as they were unable to form a clear object and angular velocity percept. Under 2D Cues, Table 6.4 lists the normalized slopes of hypothetical response curves if the angular velocity judgment had been proportional to a particular 2D image measure. 2D cues considered are mean or maximal image velocity, maximal horizontal extent of the projection, the length of the trajectory of the ellipsoid's two POBs (and fixed points), and area of the projection. In all those cases, response curves were virtually straight lines (RMS of fitted line and actual data points was less than .03, and p_F was vanishingly small).

I will consider these cues in turn.

(ii) *Mean velocity.* Velocity is defined as displacement in the horizontal direction per view, and mean velocity is obtained by averaging over all points along the contour and over all views. This results in a slope of -.44, which is steeper than the observed average of -.18 for JB and JT, and of -.32 for JL and VB (averaged over all conditions), although there are some occasions in which the observed slope matches the one based on mean velocity (JL, contour, 4°/view).

A major caveat must be made regarding the definition of image velocity. Recall that the contours are smooth ellipses without any identifiable feature points; there is no real reason

why velocity should be measured in the horizontal direction. Indeed, sometimes contours look like pulsing blobs with motion in the vertical direction.

(iii) *Maximal velocity* is defined as the maximal velocity anywhere along the contour at any time. It leads to a slope of -.43, i.e., essentially the same slope as for mean velocity.

(iv) *Maximal extent* is defined as the maximal horizontal size of the ellipsoid's projection at any time. It leads to a slope of -.17 which is too shallow.

(v) *Area* of the ellipsoid's projection summed over all views. It leads to a slope of -.12 which is too shallow.

(vi) *POB trajectories* defined as the maximal horizontal excursion of the POBs. This leads to a slope of -.31 which closely matches the observed average slope of -.32 of subjects VB and JL and the slope of subjects JB and JT with the contours; it also agrees with the fact that subjects reported tracking and somehow basing their judgments on POBs. Judgments can be based on POBs in at least the following two ways. First, judgment is proportional to the mean velocity of the POBs. This would account for the observed slopes for JL and VB, but not for the fact that the judged difference between 3°/view and 4°/view is less than 1°. In fact the mean difference for JL and VB between the two actual angular velocities averaged over the Contour and Silhouette conditions and over thickness is only $.59 \pm .36°$; the corresponding value for JB and JT is $.68 \pm .42°$ (the difference between the two values is not significant). The mean difference averaged over all subjects and conditions is $.64 \pm .39°$. This suggests that the judgments of at least subjects JL and VB might be influenced by the actual length of the POBs' trajectories which is the same for the 3°/view and the 4°/view conditions.

To summarize, equating the angular velocity of two different objects is hard; subjects did not feel confident about the accuracy or reproducibility of their judgments. The judgments of the two subjects who did not perceive solid ellipsoids decrease with thickness at a rate that can be accounted for if we assume that they use POBs. The judgments of the two subjects who did perceive solid objects also decreased with thickness but much less and in some instances not at all. I conclude that it is possible to perceive angular velocity accurately using POBs (fixed points), but the judgment is easily influenced by 2D factors such as mean image velocity. It is noteworthy that even though angular velocity judgments are not correct for subjects JB and JT, their ability to discriminate flat ellipse from solid ellipsoids is good; this insensitivity of the shape judgments to perceived angular velocity might be accounted for by the fact that the various measures of solidity do not vary much for $T > .3$ (Figure 6.10).

The above results agree with those of Kaiser (1990), the first published study on the human perception of angular velocity. The decrease in perceived angular velocity with either decreasing area, mean image velocity, or mean velocity of the POBs agrees with Kaiser's finding that a large object appears to be rotating slower than a small object. Specifically, two cubes, one half as large as the other (scaled by .5), appear to be rotating equally fast if the smaller one rotates 18.5% faster, which I interpret as meaning that the smaller cube appears to rotate 18.5% slower. In the present experiment, the decrease in perceived angular velocity ranged from 6 to 46% (for completeness I mention that Kaiser's objects rotated at .33 Hz, whereas mine rotated at .09 Hz for $\theta = 3°$/view, and .11 Hz for $\theta = 4°$/view). This decrease is accompanied by a relatively small decrease in area of only 12% as compared to the

75% decrease in Kaiser's cubes. The decrease in image velocity (mean or maximal), however, is comparable: 44% in the present study and 50% in Kaiser's. Of course, the decrease in image velocity in Kaiser's experiment is accompanied by a large decrease in area, which is not true for the present experiment.

One comment on Kaiser's experimental design deserves mentioning. The two objects whose angular velocity was to be discriminated were shown side by side for 4 seconds; in experiments 1 and 3, the two objects were cubes whose top face was visible during the entire rotation. The objects always rotated in phase or 45° out of phase. As Kaiser herself points out, the similarity of the objects and their motion opens the door to numerous strategies to determine whether two objects are rotating equally fast, strategies that have nothing to do with angular velocity. For example, a subject could track a vertex on the top face of one cube and compare its image trajectory to that of a vertex on the other cube. Or a subject could follow a particular edge on the top face of the cubes, and compare the change in orientation of their projection. Such strategies are much less likely in the present experiment because the two objects whose angular velocity is to be compared differ in shape (one a slanted ellipsoid, the other a vertical cylinder) and motion (one rocks back and forth, the other rotates continuously; both continue their motion until the subject responds). These differences in experimental design could account for the higher variability in the present data.

Having shown that the human visual system tends to underestimate the angular velocity of rotating ellipsoids, we can go back to the preceding discrimination experiment (Experiment 6.1) and compute the values of the various discrimination measures using θ_{est} instead of θ_{act}. I used Equation 4.18 to estimate angular velocity, in the same way as I did for Figure 6.11; these estimates are listed in appendix B, Tables B.1 through B.8, column θ_{est}, together with the corresponding failure rate, column %e. The resulting psychometric curves are also shown in appendix B for the motion discrimination measure (Figure B.1b), the nonrigidity measure (Figure B.1d), the rim prediction measure (Figure B.1f), and the tangent plane prediction measure (Figure B.1h). The thresholds, slopes, and RMS errors for these psychometric curves are listed in Table 6.2 for comparison with the values obtained using the ellipsoids' actual angular velocity.

Using θ_{est} instead of θ_{act}, the fit of the psychometric curves improves somewhat. Error decreases slightly in 15 out of 32 cases and remains the same in 10 cases; only in Exp. B.8 does error increase considerably. Visual inspection of the psychometric curves confirms this somewhat closer fit.

The effect on the discrimination measures of using θ_{est} instead of θ_{act} is to consistently decrease their value for the moderately fat ellipsoids while hardly affecting them for the fat ellipsoids (compare Table B.11 with B.1, and B.12 with B.4). This effect is clearly visible in the psychometric curves: With θ_{est}, slopes increase and thresholds decrease and vary less. The decrease in variation can be quantified by considering experiments B.1 through B.5 to be minor variations of each other, and computing the ratio of standard deviation and mean for the absolute and relative thresholds in these five experiments. This ratio decreases for the nonrigidity measure (from .094 to .058 for the absolute thresholds, and from .075 to .032 for the relative thresholds), for the rim prediction measure (.208 to .158, and .141 to .112), and for the absolute threshold for the motion measure (.18 to .167). The ratio for the

tangent plane prediction measure, in contrast, increases (from .171 to .202 for the absolute threshold, and from .072 to .101 for the relative threshold), as does the ratio for the relative threshold of the motion measure (from .124 to .134).

On balance one can say that the expected effect of using θ_{est} instead of θ_{act} is present, albeit small. One might suggest actually measuring the perceived angular velocity as in Experiment 6.2 for the ellipsoids used in Experiment 6.1. However, I do not think the results will improve significantly. Moreover, there remains the fact that human observers not only underestimate the angular velocity of rotating ellipsoids but also the slant of the ellipsoids. The major axis of the ellipsoids in the two experiments made an angle of between 45° and 60° with the vertical (the axis of rotation); however it looked more like between 30° and 45° even for the flat ellipse (one would expect a slight underestimate for a fat ellipsoid as the plane of its rims makes a smaller angle with the vertical than does its major axis; for example, the rim of the ellipsoid $x^2/2 + y^2 + z^2/6. = 1$, i.e., an ellipsoid with thickness 1, slanted by 60°, makes an angle of about 45° with the vertical; see section 6.3). This agrees with the results of Experiment 5.1; in the latter experiment, the slant of rotating planar surface patches is systematically underestimated especially for slants greater than 45° (Figure 5.2).

6.3 Perception of shape from contours of rotating ellipsoids

The preceding experiments established that some subjects perceive an solid object when looking at the contours or silhouettes of a rotating ellipsoid and that they can distinguish reliably between flat ellipses and solid ellipsoids. In this section, I will further investigate the 3D percept resulting from dynamic contours. The idea is to have subjects somehow indicate what shape they perceive.

There seem to be two methods for pursuing this. In the so-called *local depth probe* approach, subjects have to indicate the perceived depth at various locations within a stimulus. Stevens & Brookes (1987) successfully used a stereo depth probe to measure perceived depth in monocularly viewed line drawings of planar and undulating surfaces. Saidpour, Braunstein, & Hoffman (1990) used a dynamic probe dot, whose depth could be changed interactively by the subject, to measure the perceived surface in "structure-from-motion" displays simulating rotating cylinders.

The main advantage of this approach is that it is completely unbiased in the sense that the subject's response is not limited in any way. The disadvantages are several. First, it is slow because it takes many local probes to get an idea of the perceived surface. Second, by attending to the probe the global percept of shape may be affected (found in comparing the size of two line segments on a rotating cylinder, Braunstein, personal communication). Third, there might not be a simple relationship between the setting of the probe (disparity in the stereo probes; simulated distance from the axis of rotation with dynamic probe dots) and its perceived depth in the context of the stimulus being probed; that is, the setting of the probe might not be an independent measure of perceived depth (Stevens & Brookes, 1988, for stereo probes in the context of monocular stimuli, and Ramachandran, Cobb, & Rogers-Ramachandran, 1988, for "structure from motion" in the context of bounding contours). Finally, the probe's perceived location on the surface might differ from the intended location (see Results and Discussion under Experiment 6.3).

The alternative is to use a *global shape model* either in the form of a set of models from which the subject can choose or in the form of a model that can be adjusted interactively by the subject. In the present context, I experimented with "inflatable" ellipsoids; that is, the subject can change the thickness of an ellipsoid from zero to some large number. The main advantage is speed and the fact that it is an object with a surface having curvature, which might be more appropriate than a single dot whose only attribute is depth. The main disadvantage of course is that there is only a restricted repertoire of shapes from which to choose. A less obvious bias is that the model might influence the way in which the stimulus is perceived, for example, by its shape, angular velocity, slant, etc. A nonrigid stimulus might become rigid because of its similarity and juxtaposition to an obviously rigid probe. The complexity of the global probe increases the probability that factors other than "shape" determine the selection of a particular model.

I decided to use the local depth probe as it is the least biased.

Experiment 6.3: Probing perceived shape inside contours with a dynamic probe dot

The purpose of this experiment is to determine whether perceived depth inside dynamic contours is in accordance with the percept of a solid object or with that of a planar object.

Methods:

Subjects. Three subjects were used in this experiment: the author (JB) and two subjects (JT and JL) naive as to the purpose of this experiment but experienced with similar psychophysical experiments.

Design. There were two independent variables, namely the thickness of the ellipsoid (4 levels: 0.0, .02, 0.4, and 0.8), and the position of the probe point (4 positions). Each probe position was shown 5 times for each ellipsoid. The total number of flat ellipses equaled the number of fat ellipsoids. The flat ellipses were partitioned into three sets of equal size, each set serving as the control for one of the ellipsoids. The slant angle of the control ellipse matched the slant angle of the rim of its corresponding ellipsoid seen from the front (i.e., when it has not rotated about the vertical). Slant angles of the control ellipses for ellipsoids with thickness .2, .4, and .8 were 56.7°, 53.5° and 47.0° (Figure 6.14). The sequence of presentation was randomized.

Stimuli. Each visual display showed a sequence of contours of an ellipsoid rotating about a vertical axis at constant angular velocity $\theta = 4°$/view. Subjects saw a sequence $v_1 v_2 ... v_{n-1} v_n v_n v_{n-1} ... v_2 v_1 v_1 v_2 ...$, corresponding to an ellipsoid rotating back and forth; n = 16 views. The range of rotation was $[-30°, 30°]$. Each view was shown for 100ms, and is then replaced immediately by the next one (OSA = 100ms. ISI is zero). The direction of rotation reversed instantaneously at v_1 and v_n in the sequence. Contours were computed assuming orthographic projection and were rendered as bright curves on a dark background.

A probe dot was included inside the contours, rotating at the same angular velocity as the ellipsoid, also about the vertical. The distance between the probe and the axis of rotation could be adjusted interactively by means of a joystick. The probe was located along the vertical at locations (0.0, 0.0), (0.0,-.2), (0.0,-.4), and (0.0,-.6). Positions below the center of the ellipsoids' projection were chosen because only there is the depth perceived with an ellipsoid of the same sign as the depth perceived with an ellipse (compare the side views of the ellipsoids in Figure 6.14 with those of the ellipses). Immediately above the

center, the two differ in the sense that the surface of the ellipsoid is perceived to be in front of the axis of rotation and that of the ellipse to be behind the axis of rotation. As we are using orthographic projection, we cannot recover the sign of the perceived depth from a dynamic probe dot. Another reason for choosing probe positions below the center is that the difference in depth between ellipses and ellipsoid is largest there. That is also the reason why the probe positions were along the major axis of the contours (they remain along the major axis for ellipses but not for ellipsoids).

Apparatus. Stimuli were displayed on an IMI-455N. Subjects viewed the display in a dark room through a viewing tube which limited the field of view to approximately 10° of visual angle. The distance between the subject's eye and the screen was .9m.

Procedure. Subjects were instructed to adjust the depth of the probe such that it appears to lie on the perceived surface. Subjects could take as long as they felt necessary. Each experimental session lasted about an hour.

Results and discussion:

Figure 6.14 shows the results for the three subjects and Table 6.5 summarizes the difference in perceived depth between each ellipsoid and its control ellipse.

Subjects all agreed that the task was easy. This subjective impression is substantiated by the very small standard deviations shown in Figure 6.14, and that the planar surface of the ellipse was interpolated quite well. Subjects did not report using 2D strategies such as the matching of image velocities; they reported putting a dot on the perceived surface. Perceived slant was around 45°. Interestingly, all subjects perceived the probe dots to be considerably lower, with respect to the center of the projection, than they actually were. For example, the probe at position (0,-.6) seemed to be located at (0,-.8). This tendency to displace the probe's position downward could account for the fact that most subjects did not put the center dot at zero depth even for the ellipse (Figure 6.14); I did not investigate whether there was a systematic difference between the ellipses and the ellipsoids.

All subjects saw the top of the figure as farther away. All subjects perceived a clear surface; JT reported seeing only flat objects; JB and JL reported seeing flat and curved objects. These subjective impressions of shape are partially confirmed by the depth judgments. Figure 6.14 shows pictorially what Table 6.5 further quantifies, namely that JT and JL perceived the ellipses and the ellipsoids as flat; only in three out of 21 instances is the difference in perceived depth in the ellipsoid significantly different from that in its control ellipse. On the other hand, the differences for JB are significant in 7 out of 12; the number of significant differences increases with ellipsoidal difference, as does the mean of the differences averaged over all positions (column mean in Table 6.5). JL's subjective impression of curved surfaces is not reflected in his depth judgments.

One problem in comparing the depth judgments of the ellipsoid with those of its control is the fact that the ellipsoid's angular velocity is underestimated (section 6.2). The actual data we obtain from the subject are displacements $dx(t)$, which we translate into relative depth judgments, z, by means of $z = dx/d\theta$ or $z \propto dx/dv$, where dx is displacement during time dt, $d\theta$ the corresponding amount of rotation, and dv the magnitude of local image velocity. I used the ellipsoid's actual angular velocity (filled circles in Figure 6.14) or its actual angular velocity divided by $1 - .31T$ (open circles in Figure 6.14), where $-.31$ is the normalized slope of the line relating judged angular velocity and thickness T, if the judgment

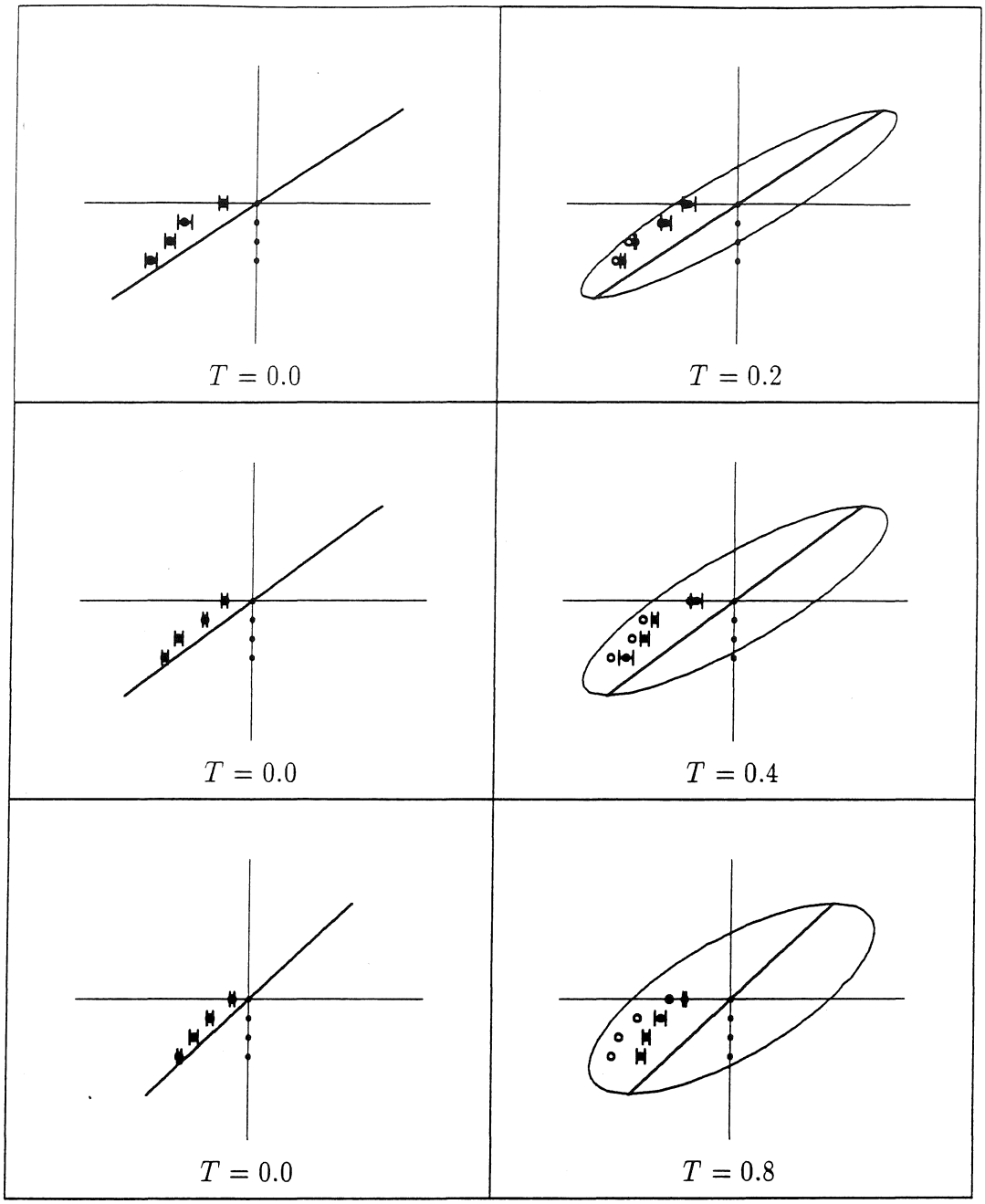


Figure 6.14a. Judgments for subject JB for three members of the family of ellipsoids $x^2/2 + y^2/T + z^2/6 = 1$. Bars indicate standard deviations (5 judgments). Open circles are judgments adjusted for lower angular velocity. Dots along the vertical indicate the vertical positions of the probe dots.

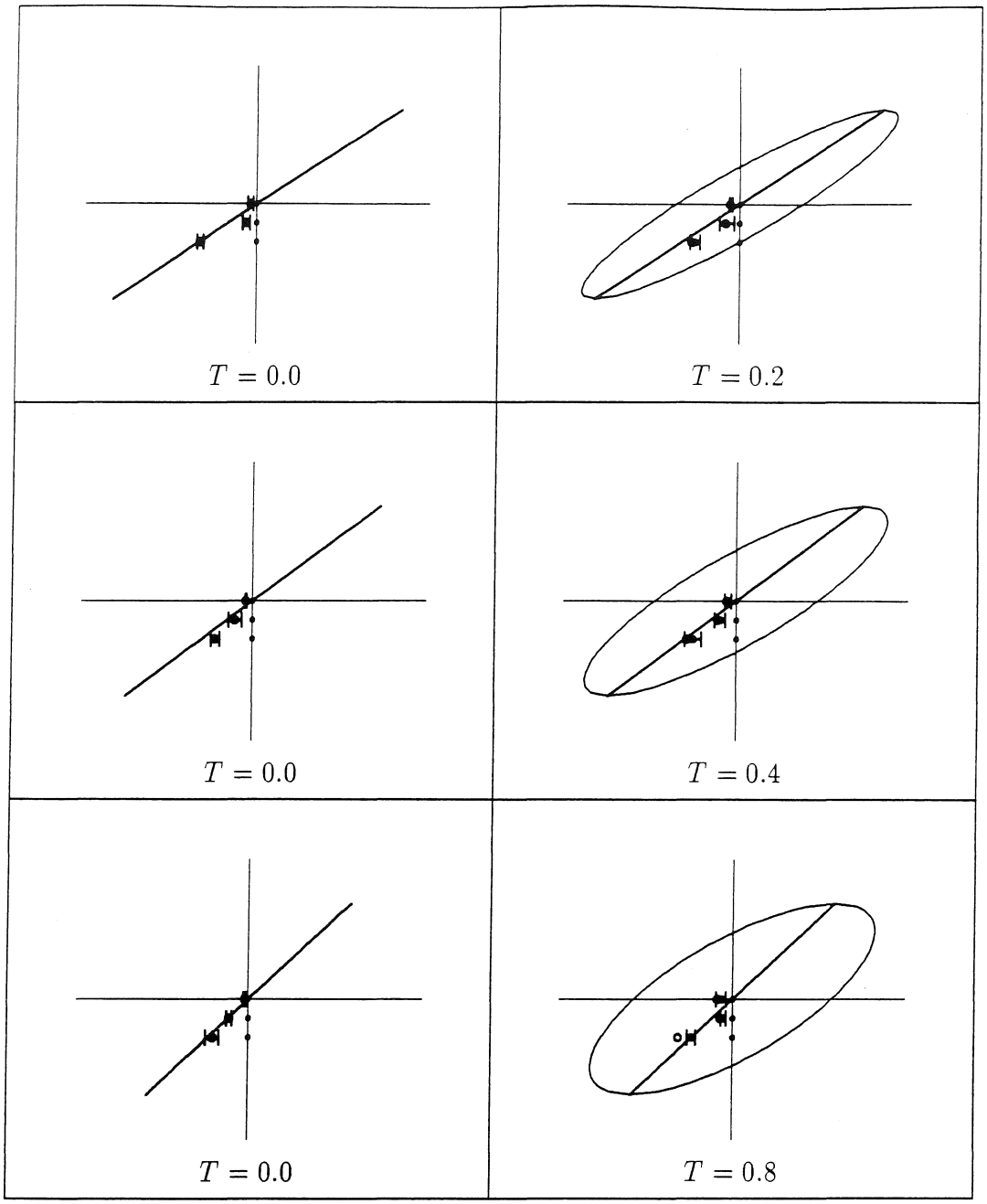


Figure 6.14b. Judgments for subject JT for three members of the family of ellipsoids $x^2/2 + y^2/T + z^2/6 = 1$. Bars indicate standard deviations (5 judgments). Open circles are judgments adjusted for lower angular velocity. Dots along the vertical indicate the vertical positions of the probe dots.

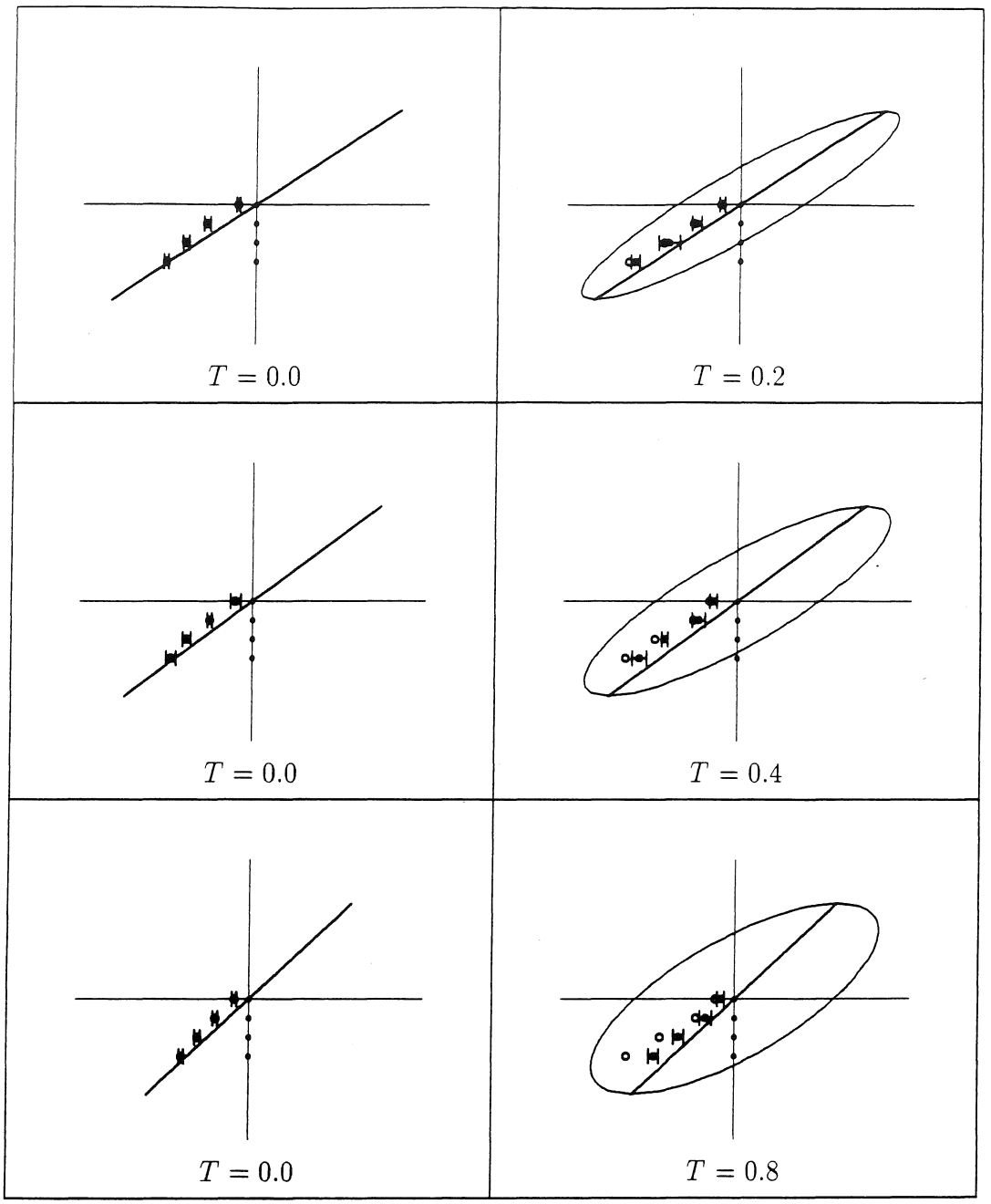


Figure 6.14c. Judgments for subject JL for three members of the family of ellipsoids $x^2/2 + y^2/T + z^2/6 = 1$. Bars indicate standard deviations (5 judgments). Open circles are judgments adjusted for lower angular velocity. Dots along the vertical indicate the vertical positions of the probe dots.

Table 6.5 Surface interpolation judgments. Values are difference between solid ellipsoid and its control. * indicates that the difference is significant at 5% level, ** at 1% level. Position is the fraction of the vertical distance between the center of an ellipsoid's projection and its lower POB. Mean is the difference between an ellipsoid and its control ellipse averaged over all positions.

Subject	Thickness	Position .0	.2	.4	.6	mean
JB	.2	.158	-.004	.160*	.092	.102
	.4	.101	.339**	.173*	.213*	.207
	.8	.312**	.337**	.312**	.221**	.296
JT	.2	.021	.018	-.135		-.032
	.4	.012	-.021	.056		.016
	.8	.081	-.111*	.052		.007
JL	.2	.010	-.056	.004	.151*	.027
	.4	.077	-.041	.071	.174	.070
	.8	.002	-.048	.053	.141*	.037

depended on the POBs trajectory (section 6.2). As the judged angular velocity is less than the actual one, the depth judgments are larger. This adjustment hardly affects JT and JL's data; the depth judgments continue to lie on a straight line. In case of JB, the difference between ellipses and ellipsoids becomes larger.

The discrepancy between JL's subjective report (50% of objects perceive as flat, 50% as curved) and his depth judgments is puzzling. Perhaps it is due to the fact that subjects actually interpolate 2D image velocities along the line connecting the top and bottom POBs. As the POBs are fixed points, the resulting depth judgments would lie on a straight line both for the ellipse and the ellipsoids. Interpolation in the horizontal direction is not as straightforward because the velocity along the contour changes over time not only due to the rotation of the ellipsoid but also due to the rim sliding over the ellipsoid's surface. Thus, positioning the probe dot on the ellipse's planar surface would be possible by interpolating image velocity in a horizontal direction and equating the probe dot's velocity with the interpolated one for that position. This would not be possible for the ellipsoid. But perhaps something like this does happen for JL which would account for the higher SD in the depth judgments for the ellipsoids (average of .110 for the ellipsoids vs. .065 for the ellipses of JL;

JB's judgments do not show this effect).

One problem with using contours as stimuli is that the perceived surface is not sustained locally by anything in the image. Results might improve by using silhouettes whose inside is brighter than the background. (This was not possible with our equipment.)

6.4 Summary

In this chapter, I studied several aspects of the human visual perception of solid shape from the occluding contours of rotating ellipsoids. I developed a psychophysical test to determine whether human subjects perceive solid shape from contours. The test requires subjects to discriminate between solid, volume-containing ellipsoids and flat ellipses. The ellipsoids vary in thickness from almost zero (virtually flat) to one third of their major axis (very fat). Two out of four subjects tested could perform this task for ellipsoids rotating about a vertical axis and the likelihood of perceiving an ellipsoid as solid increased with its thickness. They reported perceiving curved and solid objects, whereas the other two subjects perceived all stimuli as flat.

The thickness of an ellipsoid is, of course, an arbitrary measure of its solidity. More plausible measures can be obtained by basing them on explicit models for the representation of solid shape. I did this for the models outlined in section 4.3. The measures based on κ_m, κ_r, nonrigidity, or the rim prediction models could account reasonably, and equally, well for the data. The measures based on Gaussian and mean curvature could not. Furthermore, there is a 2D measure, the so-called contraction in the direction of motion, which could not be eliminated from the stimuli and which could account for some of the data.

In Chapter 4, I argued that the fixed points on the occluding contours of a rotating ellipsoid are used to compute its motion, i.e., its axis of rotation and its angular velocity. As fixed points only occur with frontoparallel axes of rotation, it follows that (i) subjects should be able to judge angular velocity and the axis of rotation for frontoparallel axes, and (ii) fail to do so for axes of rotation that are not frontoparallel, and hence be impaired in the discrimination task.

To test whether subjects correctly perceive the angular velocity of ellipsoids rotating about a vertical axis, I used ellipsoids of varying thickness and two different angular velocities. Next to the contours of these ellipsoids, there was a rotating cylinder whose angular velocity could be adjusted interactively by the subjects. The ellipsoids rotated back and forth whereas the cylinder rotated continuously in the same direction. Subjects had to equate the angular velocity of the cylinder with that of the ellipsoid or ellipse perceived from the occluding contours; they also had to indicate whether they perceived a solid or a flat object. The judgment of angular velocity proved to be very hard. Even for control stimuli consisting of dots on an ellipsoidal surface, judgments varied considerably for each subject and as much as a factor of two across subjects. Such a range would not be found for say judgments of linear velocity. The two subjects who could discriminate solid and flat objects were better at judging angular velocity than the two subjects who could not. The latter's judgment was simply proportional to the mean image velocity in the neighborhood of the fixed points.

To test whether subjects can perceive solid shape from the contours of ellipsoids rotating about a slanted axis, I determined whether the two subjects, who could accurately discriminate ellipsoids from ellipses rotating about the vertical, would continue to do so if the axis of rotation is slanted. The answer is no if the axis of rotation is slanted away from the major

axis of the ellipsoids, and yes if it is slanted toward the major axis. This difference is due to the fact that flat ellipses look like solid ellipsoids in the former but not in the latter case. Subjects had no idea where the axis of rotation was. One subject, JT, perceived all objects to be rotating about a frontoparallel axis. This hints at an interesting interpretation only possible under perspective analysis of the visual stimuli, namely that changes in vertical size of the projection are not attributed to the axis of rotation being slanted but to receding and approaching movements of an object rotating about a frontoparallel axis (or rather an axis perpendicular to the prevailing viewing direction). Information about motion in depth is of course not available under orthographic projection. However, subjects did not spontaneously report motion in depth.

These results are consistent with the hypothesis that fixed points are necessary and sufficient to perceive the motion of a rotating ellipsoid from its occluding contours.

To obtain further evidence for the perception of solid shape from occluding contours, I had subjects adjust a dynamic probe dot, located inside the occluding contours, such that it appeared to lie on the perceived surface. Subjects considered this an easy task and proved to be very accurate. Only for one of the three subjects was there a significant difference in dot placement between ellipses and ellipsoids, a difference consistent with his subjective impression of shape. One subject only perceived flat objects and his dot placements were consistent with that. And although the third subject perceived solid and flat objects, his dot placements were consistent with a perception of flat objects only. The results of this experiment might improve with silhouettes in which the figure is lighter than the background, that is, in which the figure is better able to sustain the perception of a surface. A more fundamental problem inherent in this task is that subjects have to focus their attention on a small dot, which might affect their global percept of shape at the time of dot placement.

Chapter 7
Summary and discussion

7.1 Visual perception of solid shape and object motion from occluding contours

The occluding contours of objects are prominent and highly informative features of a visual image. By circumscribing an object's projection, they divide an image into figure and ground and can delineate parts within the object (section 2.1). Observers moving through an environment with objects at various distances can compute their motion from the optic flow field in the neighborhood of occluding contours (Longuet-Higgins & Prazdny, 1980; Warren & Hannon, 1990; section 2.5). The contour of an object conveys qualitative information about the object's shape (Koenderink, 1984b; Brady, Ponce, & Yuille, 1985); if the object moves relative to the observer, quantitative shape information can be derived from its contours (Giblin & Weiss, 1987; Blake & Cipolla, 1989; section 2.7).

Psychophysical studies have shown that human observers can indeed perceive solid shape from the contours of objects rotating about a frontoparallel axis (Todd, 1985; section 2.7), and perceive an ordering in depth from moving contours and the accompanying accretion and deletion of image detail (Kaplan, 1969; Craton & Yonas, 1990; section 2.6).

I extended these studies of the human perception of solid shape from occluding contours. I focused on the situation of a stationary observer whose image consists solely of the occluding contours of a rotating solid object under orthographic projection. I inquired whether the object's motion and shape can be computed from its occluding contours, and whether the human visual system actually performs that computation.

I proposed a three-stage model for the perception of solid shape from the contours of a rotating object. In the first stage, the axis of rotation and angular velocity of the object are computed. This knowledge makes it possible to compute relative depth and various measures of surface curvature along the rim in the second stage. These local shape descriptors are then integrated, in the third and final stage, to yield a global percept of solid shape.

Using points of correspondence and orthographic projection, the object's rotation can only be computed if the axis of rotation is frontoparallel. If the axis is slanted out of the image plane, the points of correspondence are no longer the image of fixed points on the object and hence cannot be used to compute the object's motion. Under perspective projection, the axis of rotation could in principle be computed from two successive views having at least five points of correspondence (Faugeras & Maybank, 1989); it is doubtful, however, whether this would be practical. It is not possible to compute object motion from the contours themselves, that is, without using points of correspondence. Points of correspondence that are images of fixed points lie on straight lines in the image and vice versa; these lines are part of the envelope of the contours and hence easy to find.

I found that two out of the four subjects tested perceived the correct angular velocity of ellipsoids rotating about a frontoparallel axis. The same two subjects were also able to discriminate between solid ellipsoids and flat ellipses on the basis of their occluding contours. The other two subjects perceived flat objects and their angular velocity judgments were

proportional to the image velocity at the points of correspondence. Certain slants of the axis of rotation made it impossible to discriminate between solid and flat ellipsoids and to perceive the axis of rotation.

These results are consistent with the hypothesis that the human visual system uses fixed points to compute object motion, at least under my experimental conditions. The hypothesis predicts that if the the fixed points were no longer available, for example because of occlusion, subjects will no longer be able to accurately judge angular velocity and, hence, no longer be able to discriminate between solid ellipsoids and ellipses, at least not as well. The latter qualification is necessary because subjects could sometimes discriminate the two while underestimating angular velocity by as much as $1°/s$. As all subjects who participated in the angular velocity experiment reported focusing their attention on the neighborhood of the fixed points, I expect that performance will deteriorate markedly if the fixed points are occluded.

The use of fixed points would give an interesting twist to Todd's (1985) contention that correspondence is not necessary to perceive solid shape. From the fact that subjects can perceive solid shape when looking at the silhouette of a blimplike object rotating about a frontoparallel axis, Todd concluded that the human visual system does not require correspondence. However, there were two points of correspondence in his displays, points which can be used to compute the blimp's angular velocity, which in turn is needed to compute relative depth and surface curvature along the rims. However, curvature computations proved relatively insensitive to the exact angular velocity (section 6.2). Hence, it might be that points of correspondence are used only to determine the axis of rotation and the velocity field along the contours.

To test the hypothesis that the human visual system uses a perspective analysis to compute a slanted axis of rotation, we need to use objects which have at least five points of correspondence. One possibility is to interpenetrate a number of ellipsoids; the silhouette of the resulting object is the union of the silhouettes of the individual ellipsoids (section A.4 gives the algorithm for computing contours of ellipsoids under perspective projection).

Giblin & Weiss (1987) were the first to show how surface curvature could be computed from the contours of an object rotating about a frontoparallel axis. I used a different method to derive the same relationship between surface curvature in the direction of motion and the velocity field along the contour (section 4.2).

I formulated three models for the perception of solid shape (section 4.3). In the rim prediction model, relative depth along the contour is used to predict subsequent rims or contours. If predicted and actual visual input agree, the observer perceives a rigid object in the form of a 3D wire. If the two disagree, as they would for the occluding contours of a smooth object, the observer either perceives a nonrigid wire or a rigid, solid object whose internal structure derives from past views. The tangent plane prediction model is similar to the rim prediction model but uses the tangent plane as well as relative depth. In the curvature model, surface curvature is computed directly from the deforming contour and propagated inward to obtain a global curvature interpretation.

I used these models to compute how well subjects would be able to discriminate between solid ellipsoids and flat ellipses. The rim prediction model, the model based on curvature in the direction of motion, and the radial curvature model could account reasonably and equally

well of the data. The Gaussian curvature model and the mean curvature model could not account for the data. Computer simulations showed that the discrimination measures based on the rim prediction model, the radial curvature (as well as the model using curvature in the direction of motion), and the nonrigidity model were quite insensitive to angular velocity for ellipsoids that exceeded a certain thickness (section 6.2). Underestimating angular velocity for the thinner ellipsoids led to values of the discrimination measures comparable to those for flat ellipses.

There is some psychophysical evidence that the human visual system only combines pairs of successive views to compute spatial structure (Todd & Brennan, 1990). In Chapter 6, I have shown that some human observers can perceive solid shape from occluding contours, and discriminate between ellipsoids and ellipses. This is impossible on the basis of just the instantaneous velocity field because the velocity field along the contour of a rotating object merely indicates the relative depth along that contour's pre-image. In other words, there is no information about surface curvature. And because ellipsoids always have planar rims, the instantaneous velocity field of a solid ellipsoid is the same as that of a planar ellipse, making it impossible to discriminate between ellipsoids and ellipses on the basis of just two nearby views. I should point out that this a theoretical argument; it remains to be shown experimentally that human observers cannot discriminate between ellipsoids and ellipses with just two views. For the sake of argument, I will assume that they in fact cannot. How, then, can we reconcile this with Todd & Brennan's seemingly contradictory claim that the human visual system never goes beyond two views?

One possibility is to assert that Todd & Brennan's claim is not general and does not apply to occluding contours. This is not unreasonable because there at least two important differences between occluding contours and their stimuli, stimuli which consisted of apparent motion sequences of 3D configurations of between 3 and 23 points, or of 3D configurations of a few line segments. The first difference is that there is no correspondence between successive occluding contours (except for a few isolated points) and complete correspondence for the points and lines. The second difference concerns the image velocity field, which is smooth along the contours and not smooth for the points and lines. It might be that the human visual system can only compute second-order temporal relations if there is a well-defined, smooth velocity field. This hints at the problem of drawing general conclusions about the functioning of the human visual system on the basis of studies using minimal numbers of points and views (e.g., Braunstein, Hoffman, & Pollick, 1990; Todd & Brennan, 1990). There is no a priori reason to believe that what the visual system does with a handful of points has any bearing on the way it analyzes rich visual stimuli consisting of smooth velocity fields with isolated line discontinuities along occluding contours.

The other possibility is to refine Todd & Brennan's claim by saying that the human visual system cannot compute second-order temporal relations but that it can integrate three or more views. I will illustrate this distinction using the curvature models and the prediction models formulated in sections 4.3 and 6.1. The curvature models all assume that the human visual system computes curvature in the direction of motion, κ_m, via $\kappa_m = x + \partial^2 x/\partial\theta^2$. The computation of $\partial^2 x/\partial\theta^2$ is a true second-order temporal computation, requiring the time derivative of the velocity field. The prediction models all infer relative depth along the rim from the instantaneous velocity field. This information, together with the current

estimate of the object's motion, yields expectations for future visual inputs or relative depth measurements. Thus although the curvature and the prediction models both require more than two views, only the curvature model uses second-order temporal relationships.

A related topic is the perception of angular velocity. As it is impossible to compute angular velocity from just two orthographic views, a human ability to accurately perceive angular velocity would argue against the use of only two views. The results of experiment 6.2 indicate that it is hard to judge angular velocity and that linear image velocity tends to influence the judgment, which would be consistent with the two-view limitation. A related finding is that the perceived depth of rotating spheres is inversely proportional to the slant angle of the axis of rotation (Loomis & Eby, 1988, 1989); Loomis & Eby did not mention whether subjects could perceive the slanted axis of rotation. Informal observations of rotating faceted spheres suggest to me that it is difficult to percieve the axis of rotation. Clearly, more experiments are needed on the perception of angular velocity and the axis of rotation, in their own right but also in connection with the suggested two-view limitation.

7.2 Visual perception of surface orientation from motion

Numerous psychophysical studies over the last few decades have demonstrated that the human visual system can use motion to infer 3D structure (for reviews see Braunstein, 1976, 1988; Nakayama, 1985). The theoretical accounts advanced for this capability can roughly be divided into two groups on the basis of the type of 3D structure that is computed. I will refer to them as the *(relative) depth* group, and the *depth gradient* or *slant* group.

The relative depth models assert that the 3D structure computed from a sequence of images is primarily in the form of relative depth. Slant could be computed from relative depth, but it would be secondary to depth. The most well-known proponent of this group would probably be Ullman, who showed that "given three distinct orthographic projections of four non-coplanar points in a rigid configuration, the structure and motion compatible with the three views are uniquely determined up to a reflection about the image plane" (Ullman, 1979, p. 193), where by structure is meant relative depth of the points. Of course, this statement is an expression of competence rather than performance, or, to use Marr's (1982) terminology, it is a computational theory. Based on this theorem, several algorithms for computing depth have been formulated (Ullman, 1984a; Hildreth, Grzywacz, Adelson, & Inada, 1990). Given a richly textured surface, an algorithm based on Ullman's theorem is applied to local clusters of four points. Related depth models were formulated by Bennett, Hoffman, Nicola, & Prakash (1989), Faugeras & Maybank (1989), Hoffman & Bennett (1985, 1986), Huang & Lee (1989), and Tsai & Huang (1984).

The slant models assert that 3D structure is in the form of surface slant (depth gradient) with respect to the local viewing direction. Perhaps the best-known proponents of such a model are Koenderink & van Doorn (1975; 1976b; Koenderink, 1986), who showed how the first-order differential invariants of the optic flow field are related to surface slant and observer motion (section A.6); in particular, they showed that the magnitude of the deformation is the product of the magnitude of slant (or equivalently, the tangent of the slant angle) and A_t (the projection on the optic array of the relative translational motion of the observer). Other slant models were formulated by Clocksin (1980), Hoffman (1982), Longuet-Higgins & Prazdny (1980), and Nakayama & Loomis (1974). The following comments are limited to Koenderink & van Doorn's slant model.

The relative depth and slant models are sometimes characterized and contrasted as position and velocity based models. This characterization captures *how* the system goes about computing shape, where my terminology (depth vs. slant) capture *what* the visual system computes. Whereas it is probably, but not necessarily, true that implementations of currently available depth models would use successive positions of points as their input (Grzywacz & Hildreth, 1987), it is not true that slant models would lead to systems that have to use the velocity field as input. In section A.6, I used the positions of three neighboring and non-collinear points to compute divergence, curl, and deformation. Divergence can also be computed as the relative change in image area or change in texture density. And deformation can be computed from the relative change in orientation as $\dot{\triangle\phi}/\triangle\phi = \text{def}\cos 2(\phi - \mu)$, where ϕ is the orientation of some image detail, μ is the orientation of the axis of dilation, $\triangle\phi$ is a small difference in orientation of two neighboring image elements, and $\dot{\triangle\phi}$ is the time change of that difference (Koenderink, 1985, 1986). Implementations of the slant model that do not use the velocity field and its spatial derivatives escape the criticism of Nakayama (1985), who pointed out that the human perception of velocity differences follows Weber's law; that is, the human visual system measures ratios of two stimuli rather than their differences, whereas the differential invariants are based on differences rather than ratios.

A fundamental criticism of the use of the first-order differential invariants to compute slant is that slant and A_t enter as products in the expressions relating the invariants with slant and observer motion (Equation A.20; Longuet-Higgins, 1986). Thus, the object's structure is known up to a relief transformation (scaling in depth). This ambiguity can only be resolved through information other than the differential invariants, for example, through knowledge of the observer's relative motion which can be computed from the image (section 2.5 for a discussion of the potential role of occluding contours in the computation of egomotion; Heeger & Jepson, 1989; Warren & Wertheim, 1990). On the other hand, even with the ambiguity unresolved many important questions regarding an object's shape, such as whether it is planar or not, can still be answered (Todd & Brennan, 1990). Koenderink & van Doorn (1986) point out that stereoviews of reliefs of an object are virtually indistinguishable from the object itself.

The depth and slant models differ in an interesting manner with respect to the spatial coordinate systems they use. The depth models yield relative depth in an object-centered coordinate system, whereas the slant models yield the slant field in a viewer-centered coordinate system. The situation is actually more complicated for the depth models because they typically assume that a particular point in the image is followed during the computation, making it effectively the origin of a local object-centered coordinate system, and inducing an apparent rotation about an axis through that point (I hesitate to use the term foveation because it suggests that eye movements are involved). Consequently, relative depth is expressed within this local coordinate system. The question now is how all these local depth values are to be combined (assuming they are). The reason for following points instead of imposing a single global coordinate system is that it is hard to imagine that the human visual system has access to absolute positional information over distances in the image on the order of degrees.

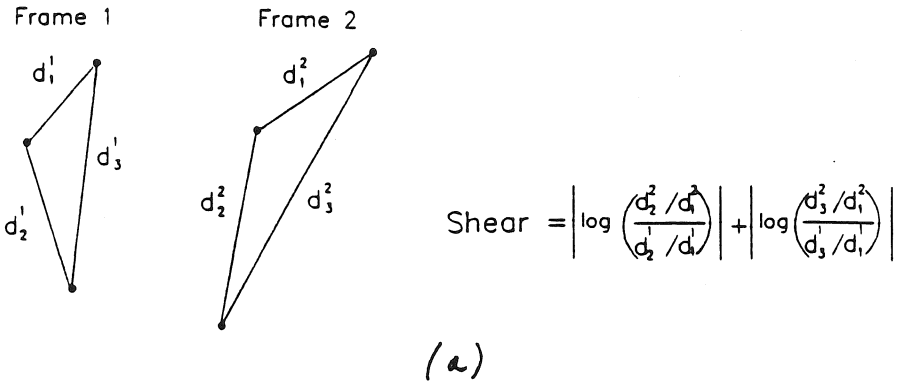
This is not to say that the depth models cannot yield slant or that the slant models cannot yield viewpoint-independent shape information. Given sufficiently many relative

depth measurements in a neighborhood it is straightforward to compute the local depth gradient with respect to the viewer. Given the slant field one could compute relative depth (see Stevens & Brookes, 1987a, for a discussion of the derivation of depth from slant in monocularly viewed line drawings). More interesting is that conjugation relationships can be computed from the slant field (section A.7 and Figure 4.2), so that we can immediately tell whether the surface is locally planar, elliptic, or hyperbolic. Assuming we know the direction of A_t, but not its magnitude, and we know the magnitude of the deformation and the axis of contraction, then we know the local slant field up to a scale factor as A_t is locally constant; thus, we can still compute the conjugation relationships even if we only know the slant field up to a scale factor.

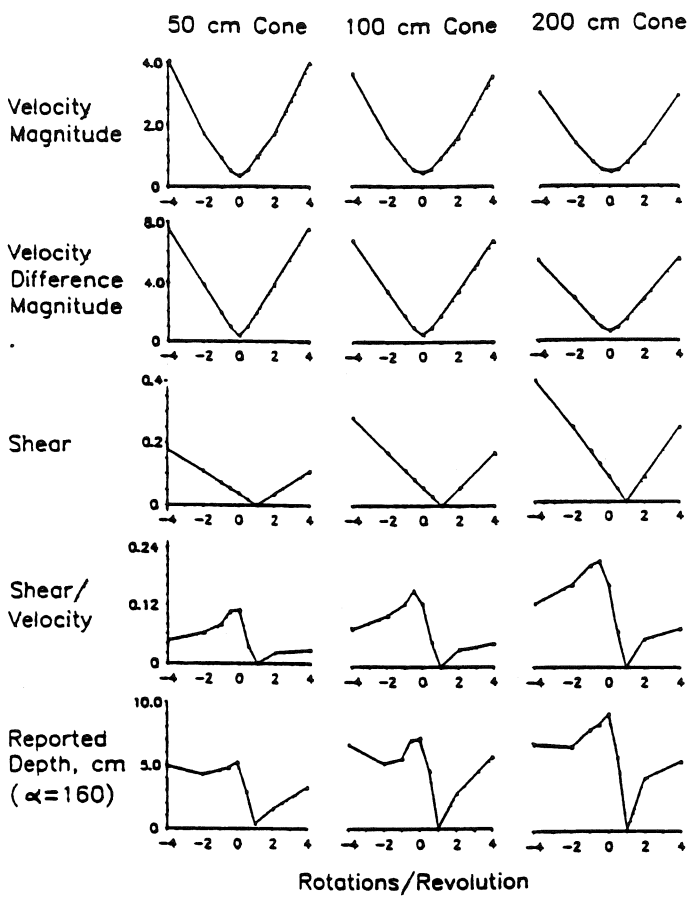
Let us now apply the depth and slant models to a concrete situation and compare their expected performance with that of human observers. Consider a richly textured, slanted plane rotating about the vertical and assume we want to judge its slant (Chapter 5). The image data consist of orthographic views of this plane as it rotates through say 20°, so that there is a well-defined image velocity field as well as a number of discrete views separated by a reasonable amount of rotation (see for example Hildreth, Grzywacz, Adelson, & Inada, 1990, for the effect of the amount of rotation on perceived structure). Applying Ullman's algorithm to this stimulus would yield the correct slant (although his structure-from-motion theorem requires that the points be non-coplanar, a problem only arises if the plane rotates about an axis obtained by intersecting the plane and the image plane, Ullman, 1979, p. 173). Applying Koenderink's slant model yields the correct slant only if we assume that the visual system can compute the plane's angular velocity.

Under the experimental conditions of Chapter 5, slant was perceived veridically up until 45° but was underestimated thereafter; in fact simulated slants between 45° and 70° were perceived as about 48°. Clearly, neither the depth nor the slant models would have predicted this result. Interestingly, perceived slant is correlated with the magnitude of the deformation per second. Deformation increases from .121/s for 22° to .176/s for 46° and is a constant .192/s for slants of 58° and 70°. Particularly interesting is that these deformations can predict the perceived slant of translating planes under perspective projection, and explain why slant is underestimated so much under these conditions (the most extreme example is that of a simulated slant of 77° with a deformation of .227/s, which is judged to be only 45°).

The only other study relating the deformation of the velocity field to the perception of shape is by Loomis & Eby (1988), who related depth judgments with what they called the shear of the velocity field. Their definition of shear is shown in Figure 7.1a; shear is not identical to the magnitude of the deformation but the two obviously covary. Their subjects looked at the perspective projection of cones whose length axis was perpendicular to the image plane; the cones rotated simultaneously about their length axis and about an axis also perpendicular to the image plane but outside the cone (the outside axis; revolution refers to rotation about the outside axis). By varying the two angular velocities different amounts of shear can be produced; if the angular velocities are equal, the cone's motion is a pure rotation in the image and shear vanishes. The cones, which pointed away from the observer and whose bases were 50cm from the observer, came in heights of 50, 100, and 200cm; the base radius of all cones was 3.5cm. Note that this implies that the conical surface



(a)



(b)

Figure 7.1. (a) Definition of shear. (b) Judged depth, shear, and shear/velocity as a function of rotations/revolutions for the three cones (from Loomis & Eby, 1988).

had a very high slant, ranging from about 80° to 90°.

Figure 7.1b shows the resulting depth judgments for a rotation of 160° about the outside axis. Estimated depth is an order of magnitude less than the actual depth and increases very little with the height of the cones. Converting the depth judgments into slant angles renders the difference between perceived and simulated surface much less dramatic (a depth of 5cm corresponds to approximately 55° slant), and could account for the small effect of cone height. The authors do not mention whether subjects perceived slant or not. The shear shown in Figure 7.1b was computed by averaging over all triplets formed by the 256 image points (it is not mentioned explicitly between what frames shear is computed but the context suggests that it is between the first and second frame, which are separated by one degree of revolution about the outside axis, and by 16ms; if this is the case I am rather puzzled by the high values for shear; it is also not clear whether shear is averaged over all triplets of image points or over all triplets of neighboring points). Interestingly, judged depth as a function of rotations/revolution is roughly similar to shear scaled by the average image velocity (bottom two rows of Figure 7.1b). The only apparent difference is between 0 and -4 rotations/revolution; judged depth has a minimum value at -2 rotations/revolutions whereas shear/velocity decreases monotonically (negative values for rotations/revolutions indicate different directions of rotation). Given that subjects had to judge depth rather than slant it is not clear how Loomis & Eby's data can be linked to the present results on slant judgments. Moreover it is not obvious how their shear is related to deformation and hence slant.

The hypothesis that judged slant is simply proportional to the magnitude of the deformation per time unit, opens the door to many experiments. For example, increasing the angular velocity of rotating planes or the linear velocity of translating planes without changing their slant should increase perceived slant because the deformation per time unit increases.

We also need to investigate the possibility that the underestimation of slant is due to the use of a *single* planar surface. Warren & Hannon (1990; see section 2.5 for a discussion of their findings) found that human observers can only perceive their direction of heading in the presence of motion discontinuities in the visual field. Perceiving the direction of heading is tantamount to decomposing the observer's motion into a translation and eye rotation. In other words, the observer is now able to factor out the rotational component of the velocity field. As A_t is retinal image motion due to translation, the observer can now convert deformation into slant via $\sigma = \text{atan}(\text{def}/A_t)$. Because neither the translating planes of Braunstein & Tittle (1988) nor the rotating planes used in Chapter 5, contained motion discontinuities, the visual system was unable to compute A_t and therefore responded on the basis of deformation per se.

7.3 Visual perception of shape from contours and optic flow fields combined

In this dissertation, I studied the human visual perception of solid shape from two separate sources, namely motion (Chapter 5) and occluding contours (Chapter 6), in isolation.

I found that occluding contours can give rise to a percept of solid shape but that it is not very compelling. Part of the reason might be that contours are generally smooth and have no any identifiable feature points, making it hard to get a well-defined velocity field. There is also nothing in the image inside the contours that would sustain the perceived surface (no motion, no differences in shading, no stereo).

Similarly, I found that the optic flow field induced by a rotating, isolated planar patch leads to a percept of slant which is veridical up to 45° slant and then remains at about that level. Translating, single planes under perspective projection also do not seem to lead to perceived slants of more than 45° even though the simulated slant can be as high as 77°. Assuming human observers can perceive slants beyond 45°, something seems to be missing from these stimuli.

What is missing from the retinal motion field induced by single planar patches is discontinuities (section 7.2), in other words, occluding contours. Occluding contours of smooth objects make at least two contributions to the motion field and the perception of shape. First, they introduce motion discontinuities thought to be important in factoring the motion field into its translational and rotational components. Second, they indicate, unequivocally, where in the image the slant angle is 90°, which might be important as a reference. Conversely, the motion field makes two contributions to the contours. First, it allows the contours to be defined in an ecologically valid manner, namely as motion discontinuities. And second, it provides a substrate to sustain whatever shape information is derived from the contours and propagated inside the figure (the substrate does not have to contain shape information; all it has to do is indicate that there is a surface; one example would be random dots moving randomly). So far it has only been demonstrated qualitatively that contours can influence the perception of shape from motion (Ramachandran, Cobb, & Rogers-Ramachandran, 1988) and shading (Ramachandran, 1988; Todd & Reichel, 1989). Indeed, Todd & Reichel (1989) found that it is hard to perceive 3D structure from shading if there are no occluding contours in the image (specifically, if the simulated surface does not have any occluding contours from the vantage point in question). This is reminiscent of Richards (1977) finding that monocular contours enhance the depth perceived with random dot stereograms. The combination of contours and motion promises to be very important for the human visual perception of solid shape.

Bibliography

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300,* 523–525.

Andersen, G. J., & Braunstein, M. L. (1983). Dynamic occlusion in the perception of rotation in depth. *Perception & Psychophysics, 34,* 356–362.

Andersen, G. J., & Cortese, J. M. (1989). 2-D contour perception resulting from kinetic occlusion. *Perception & Psychophysics, 46,* 49–55.

Arnold, V. I. (1984). *Catastrophe theory.* Berlin: Springer-Verlag.

Bennett, B. M., Hoffman, D. D., Nicola, J., & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America A, 6,* 1052–1069.

Beusmans, J. M. H. (in press). Computing occluding contours using spherical images. *Computer Vision, Graphics, and Image Processing.*

Beusmans, J. M. H., Hoffman, D. D., & Bennett, B. M. (1987). Description of solid shape and its inference from occluding contours. *Journal of the Optical Society of America A, 4,* 1155–1167.

Blake, A., & Cipolla, R. (1989). *Robust estimation of surface curvature from deformation of apparent contours.* (Tech. Rep. No. OUEL 1797/89). Oxford, UK: University of Oxford, Department of Engineering Science.

Brady, M., Ponce, J., & Yuille, A. (1985). Describing surfaces. *Computer Vision, Graphics, and Image Processing, 32,* 1–28.

Braunstein, M. L. (1968). Motion and texture as sources of slant information. *Journal of Experimental Psychology, 78,* 247–253.

Braunstein, M. L. (1976). *Depth perception through motion.* New York: Academic Press.

Braunstein, M. L. (1988). The empirical study of structure from motion. In W. N. Martin & J. K. Aggarwal (Eds.), *Motion understanding: robot and human vision* (pp. 101-142). Hingham, MA: Kluwer Academic Publishers.

Braunstein, M. L., Andersen, G. J., & Riefer, D. M. (1982). The use of occlusion to resolve ambiguity in parallel projections. *Perception & Psychophysics, 31,* 261–267.

Braunstein, M. L., Hoffman, D. D., & Pollick, F. E. (1989). Discriminating rigid from nonrigid motion: minimum points and views. *Perception & Psychophysics, 47,* 205–211.

Braunstein, M. L., & Tittle, J. S. (1988). The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 582–590.

Bruce, J. W. (1984a). Motion pictures: an application of singularity theory. *Journal of London Mathematical Society, 30,* 160–170.

Bruce, J. W. (1984b). Seeing—the mathematical viewpoint. *The mathematical intelligencer, 6,* 4, 18–25.

Bruce, J. W., & Giblin, P. J. (1984). *Curves and singularities.* Cambridge, UK: Cambridge University Press.

Callahan, J., & Weiss, R. (1985). A model for describing surface shape. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 240–245). Washington, DC: IEEE Computer Society Press.

Clocksin, W. F. (1980). Perception of surface slant and edge labels from optical flow: a computational approach. *Perception, 9,* 253–269.

Cornsweet, T. N. (1970). *Visual perception.* Orlando, Florida: Academic Press.

Craton, L. G., & Yonas, A. (1990). Kinetic occlusion: Further studies of the boundary-flow cue. *Perception & Psychophysics, 47,* 169–179.

Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General, 113,* 198–216.

Dufour, J. P. (1983). Familles de courbes planes differentiables. *Topology, 22,* 449–474.

Falmagne, J.-C. (1985). *Elements of psychophysical theory.* New York: Oxford University Press.

Faugeras, O. D. (1990). *On the motion of 3D curves and its relationship to optical flow.* (Tech. Rep. No. 1183). Le Chesnay, France: Institut National de Recherche en Informatique et en Automatique.

Faugeras, O. D., & Maybank, S. (1989). Motion from point matches: multiplicity of solutions. In *Proceedings of the IEEE Workshop on Visual Motion* (pp. 248–255). Washington, DC: IEEE Computer Society Press.

Faux, I. D., & Pratt, M. J. (1979). *Computational geometry for design and manufacture.* New York: John Wiley & Sons.

Giblin, P. J., & Weiss, R. (1987). Reconstruction of surfaces from profiles. In *Proceedings of the First International Conference on Computer Vision* (pp. 136–144). Washington, DC: IEEE Computer Society Press.

Giblin, P. J., & Soares, M. G. (1988). On the geometry of a surface and singular profiles. *Image and Vision Computing, 6,* 225–234.

Gibson, J. J., Kaplan, G. A., Reynolds, H. N., Jr., & Wheeler, K. (1969). The change from visible to invisible: a study of optical transitions. *Perception & Psychophysics, 5,* 113–116.

Gibson, J. J., Olum, P., & Rosenblatt, F. (1955). Parallax and perspective during aircraft landings. *American Journal of Psychology, 68,* 372–385.

Gigus, Z., & Malik, J. (1990). Computing the aspect graph for line drawings of polyhedral objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12,* 113–122.

Gillam, B. & Borsting, E. (1988). The role of monocular regions in stereoscopic displays. *Perception, 17,* 603–608.

Grzywacz, N. M., & Hildreth, E. C. (1987). Incremental rigidity scheme for recovering structure from motion: position-based versus velocity-based formulations. *Journal of the Optical Society of America A, 4,* 503–518.

Gualtieri, J. A., Baugher, S., & Werman, M. (1989). The visual potential: one convex polygon. *Computer Vision, Graphics, and Image Processing, 46,* 96–130.

Hebert, M., & Kanade, T. (1985). The 3D-profile method for object recognition. In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 458–463). Washington, DC: IEEE Computer Society Press.

Heeger, D. J., & Jepson, A. (1989). *Visual perception of three-dimensional motion.* (Tech. Rep. No. 124). Cambridge: MIT Media Laboratory.

Hilbert, D., & Cohn-Vossen, S. (1983). *Geometry and the imagination.* New York: Chelsea Publishing Company.

Hildreth, E. C. (1984). *The measurement of visual motion.* Cambridge, MA: The MIT Press.

Hildreth, E. C., Grzywacz, N. M., Adelson, E. H., & Inada, V. K. (1990). The perceptual buildup of three-dimensional structure from motion. *Perception & Psychophysics, 48,* 19–36.

Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America, 72,* 888–892.

Hoffman, D. D. & Bennett, B. M. (1985). Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America A, 2,* 350–353.

Hoffman, D. D. & Bennett, B. M. (1986). The computation of structure from fixed-axis motion: rigid structures. *Biological Cybernetics, 54,* 71–83.

Horn, B. K. P. (1984). Extended Gaussian images. *Proceedings of the IEEE, 72,* 1671–1686.

Huang, T. S., & Lee, C. H. (1989). Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11,* 536–540.

Husain, M., Treue, S., & Andersen, R. A. (1989). Surface interpolation in three-dimensional structure-from-motion perception. *Neural Computation, 1,* 324–333.

Ikeuchi, K., & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence, 17,* 141–184.

Julesz, B. (1964). Binocular depth perception without familiarity cues. *Science, 145,* 356–362.

Kaiser, M. K. (1990). Angular velocity discrimination. *Perception & Psychophysics, 47,* 149–156.

Kaiser, M. K., Perrone, J. A., Andersen, J. A., Lappin, J. L., & Proffitt, D. R. (1990). The effect of motion on surface slant perception. *Investigative Ophthalmology and Visual Science, Supplement, 31,* 523.

Kaplan, G. A. (1969). Kinetic disruption of optical texture: The perception of depth at an edge. *Perception & Psychophysics, 6,* 193–198.

Koenderink, J. J. (1984a). The internal representation of solid shape and visual exploration. In L. Spillmann & B. R. Wooten (Eds.), *Sensory experience, adaptation, and perception* (pp. 123–142). Hillsdale, NJ: Lawrence Erlbaum Associates.

Koenderink, J. J. (1984b). What does the occluding contour tell us about solid shape? *Perception, 13,* 321–330.

Koenderink, J. J. (1985). Space, form, and optical deformations. In D. J. Ingle, M. Jeannerod, & D. N. Lee (Eds.), *Brain mechanisms and spatial vision* (NATO ASI Series D, No. 21, pp. 31–58). Dordrecht, The Netherlands: Nijhoff Publishers.

Koenderink, J. J. (1986). Optic flow. *Vision Research, 26,* 161–180.

Koenderink, J. J. (1987). An internal representation for solid shape based on the topological properties of the apparent contour. In W. Richards & S. Ullman (Eds.), *Image Understanding 1985–1986* (pp. 257–285). Norwood, NJ: Ablex Publ.

Koenderink, J. J. (1990). *Solid shape.* Cambridge, Mass.: The MIT Press.

Koenderink, J. J., & van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta, 22,* 773–791.

Koenderink, J. J., & van Doorn, A. J. (1976a). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics, 21,* 29–35.

Koenderink, J. J., & van Doorn, A. J. (1976b). Local structure of movement parallax of the plane. *Journal of the Optical Society of America, 66,* 717–723.

Koenderink, J. J., & van Doorn, A. J. (1976c). The singularities of the visual mapping. *Biological Cybernetics, 24,* 51–59.

Koenderink, J. J., & van Doorn, A. J. (1976d). Visual perception of rigidity of solid shape. *Journal of Mathematical Biology, 3,* 79–85.

Koenderink, J. J., & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics, 32,* 211-216.

Koenderink, J. J., & van Doorn, A. J. (1981). Exterospecific component of the motion parallax field. *Journal of the Optical Society of America, 71,* 953–957.

Koenderink, J. J., & van Doorn, A. J. (1982a). The shape of smooth objects and the way contours end. *Perception, 11,* 129–137.

Koenderink, J. J., & van Doorn, A. J. (1982b). Perception of solid shape and spatial lay-out through photometric invariants. In R. Trappl (Ed.), *Cybernetics and Systems Research,* North-Holland Publishing Co.

Koenderink, J. J., & van Doorn, A. J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America A, 3,* 242–249.

Koenderink, J. J., & van Doorn, A. J. (1987). Facts on optic flow. *Biological Cybernetics, 56,* 247–254.

Korn, M. R., & Dyer, C. R. (1987). 3-D multiview object representation for model-based object recognition. *Pattern Recognition, 20,* 91–103.

Lee, C.-H., & Rosenfeld, A. (1985). Improved methods of estimating shape from shading using the light source coordinate system. *Artificial Intelligence, 26,* 125–143.

Lipschutz, M. M. (1969). *Differential geometry.* New York: McGraw-Hill.

Little, J. J. (1983). An iterative method for reconstructing convex polyhedra from extended Gaussian images. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 247–250). Los Altos, California: William Kaufmann.

Longuet-Higgins, H. C. (1986). Visual motion ambiguity. *Vision Research, 26,* 181–183.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of Royal Society of London B, 208,* 385–397.

Loomis, J. M., & Eby, D. W. (1988). Perceiving structure from motion: Failure of shape constancy. In *Proceedings of the Second International Conference on Computer Vision* (pp. 383–391). Washington, DC: IEEE Computer Society Press.

Loomis, J. M., & Eby, D. W. (1989). Relative motion parallax and the perception of structure from motion. In *Proceedings IEEE Workshop on Visual Motion* (pp. 204–211). Washington, DC: IEEE Computer Society Press.

Marascuilo, L. A. (1970). Extensions of the significance test for one-parameter signal detection hypotheses. *Psychometrika, 35,* 237–243.

Marr, D. (1982). *Vision.* San Francisco, CA: W. H. Freeman.

Mitchison, G. J., & McKee, S. P. (1985). Interpolation in stereoscopic matching. *Nature, 315,* 402–404.

Mitchison, G. J., & McKee, S. P. (1987). Interpolation and the detection of fine structure in stereoscopic matching. *Vision Research, 27,* 295–302.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern recognition mechanisms* (pp. 117–151). Rome: Vatican Press.

Mutch, K. M., & Thompson, W. B. (1985). Analysis of accretion and deletion at boundaries in dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 7,* 133-138.

Nakayama, K. (1985). Biological image motion processing: a review. *Vision Research, 25,* 625–660.

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis. *Perception, 3,* 63–80.

Nakayama, K., & Silverman, G. H. (1988). The aperture problem–II. Spatial integration of velocity information along contours. *Vision Research, 28,* 747–753.

Nelson, J. I. (1986). Unsolved problems in the cellular basis of stereopsis. In J. D. Pettigrew, K. J. Sanderson, & W. R. Levick (Eds.), *Visual Neuroscience* (pp. 405–420). Cambridge, UK: Cambridge Univ. Press.

Ninio, J. (1985). Orientational versus horizontal disparity in the stereoscopic appreciation of slant. *Perception, 14,* 305–314.

Pentland, A. P. (1982). Finding the illuminant direction. *Journal of the Optical Society of America72,* 448–455.

Perrone, J. A. (1982). Visual slant underestimation: a general model. *Perception, 11,* 641–654.

Perrett, D. I., & Harries, M. H. (1988). Characteristic views and the visual inspection of simple faceted and smooth objects: 'tetrahedra and potatoes.' *Perception 17,* 703–720.

Plantinga, H., & Dyer, C. (1987). *The asp: A continuous viewer-centered representation for 3D object recognition* (Tech. Rep. No. 682). Madison: University of Wisconsin, Computer Sciences Department.

Pollick, F. E. (1989). Shape perception from dynamic occluding contours. *Investigative Ophthalmology and Visual Science, Supplement, 30,* 264.

Ponce, J., & Kriegman, D. J. (1990). Computing exact aspect graphs of curved objects: parametric surfaces. In *Proceedings 8th International Joint Conference on Artificial Intelligence* (pp. 1074–1079). Cambridge, Mass.: MIT Press.

Prazdny, K. (1983). Stereoscopic matching, eye, position,and absolute depth. *Perception, 12,* 151–160.

Proffitt, D. R., Bertenthal, B. I., & Roberts, R. J., Jr. (1984). The role of occlusion in reducing multistability in moving point-light displays.*Perception & Psychophysics, 36,* 315–323.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331,* 163–166.

Ramachandran, V. S., Cobb, S., & Rogers-Ramachandran, D. (1988). Perception of 3D structure from motion: The role of velocity gradients and segmentation boundaries. *Perception & Psychophysics, 44,* 390–393.

Ramachandran, V. S., Madhusudhan Rao, V. & Vidyasagar, T. R. (1973). Apparent motion with subjective contours. *Vision Research, 13,* 1399–1401.

Ratliff, F. (1972). Contour and contrast. *Scientific American, 226, 6,* 91–101.

Reichardt, W., Schlögl, R, W., & Egelhaaf, M. (1988). Movement detectors provide sufficient information for local computation of 2-D velocity field. *Naturwissenschaften, 75,* 313–315.

Richards, W. A. (1977). Stereopsis with and without monocular contours. *Vision Research, 17,* 967–969.

Richards, W. A., Koenderink, J. J., & Hoffman, D. D. (1987). Inferring three-dimensional shapes from two-dimensional silhouettes. *Journal of the Optical Society of America A, 4,* 1168–1175.

Rieger, J. H. (1983). Information in optical flows induced by curved paths of observation. *Journal of the Optical Society of America, 73,* 339–344.

Rieger, J. H. (1986). Three-dimensional motion from fixed points of a deforming profile curve. *Optics Letters, 11,* 123–125.

Rieger, J. H. (1987). On the classification of views of piecewise smooth objects. *Image and Vision Computing, 5,* 91–97.

Rieger, J. H. & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A, 2,* 354–360).

Rogers, B., & Graham, M. E. (1983). Anisotropies in the perception of three-dimensional surfaces. *Nature, 221,* 1409–1411.

Rogers, B., & Koenderink, J. (1986). Monocular aniseikonia: a motion parallax analogue of the disparity-induced effect. *Nature, 322,* 62–63.

Rosenfeld, A. (1986). *Recognizing unexpected objects: a proposed approach* (Tech. Rep. No. CAR-TR-243). College Park: University of Maryland, Center for Automation Research.

Rosinski, R. R., & Levine, N. P. (1976). Texture gradient effectiveness in the perception of surface slant. *Journal of Experimental Child Psychology, 22,* 261–271.

Saidpour, A., Braunstein, M. L., & Hoffman, D. D. (1990). Interpolation in structure from motion. *Investigative Ophthalmology and Visual Science, Supplement, 31,* 522.

Seales, W. B., & Dyer, C. R. (1990). *Modeling the rim appearance* (Tech. Rep. No. 931). Madison: University of Wisconsin, Computer Sciences Department.

Shimojo, S., Silverman, G. H., & Nakayama, K. (1988). An occlusion-related mechanism of depth perception based on motion and interocular sequence. *Nature, 333,* 265–268.

Stevens, K. A., & Brookes, A. (1987a). Depth reconstruction in stereopsis. In *Proceedings First International Conference on Computer Vision* (pp. 682–686). Washington, DC: IEEE Computer Society Press.

Stevens, K. A., & Brookes, A. (1987b). Probing depth in monocular images. *Biological Cybernetics, 56,* 355–366.

Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research, 28,* 371–386.

Stewman, J. H., & Bowyer, K. W. (1987). Aspect graphs for convex planar-face objects. In *Proceedings IEEE Computer Society Workshop on Computer Vision* (pp. 123–130). Washington, DC: IEEE Computer Society Press.

Takasaki, H. (1970). Moiré topography. *Applied Optics, 9,* 1467–1472.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233–282.

Thompson, W. B., Mutch, K. M., & Berzins, V. A. (1985). Dynamic occlusion analysis in optical flow fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 7,* 374–383.

Thompson, W. B. (1989). Structure-from-motion by tracking occlusion boundaries. *Biological Cybernetics, 62,* 113–116.

Todd, J. T. (1985). Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance, 11,* 689–710.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional structure from minimal apparent motion sequences. *Perception & Psychophysics, 48,* 419-430.

Todd, J. T., & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review, 96,* 643–657.

Tsai, R. Y., & Huang, T. S. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 6,* 13–37.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, Mass.: MIT Press.

Ullman, S. (1984a). Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception, 13,* 255–274.

Ullman, S. (1984b). Visual routines. *Cognition, 18,* 97–159.

Van Doorn, A. J., & Koenderink, J. J. (1982). Visibility of movement gradients. *Biological Cybernetics, 44,* 167–175.

Van Doorn, A. J., & Koenderink, J. J. (1983). Detectability of velocity gradients in moving random-dot patterns. *Vision Research, 23,* 799–804.

Van Hove, P. (1987). Silhouette-slice theorems. In *Proceedings of the IEEE Computer Society Workshop on Computer Vision* (pp. 295–297). Washington, DC: IEEE Computer Society Press.

Wallach, H. & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45,* 205–217.

Warren, R. & Wertheim, A. H. (Eds.). (1990). *Perception & control of self-motion.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Warren, W. H., Morris, M. W., & Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 646–660.

Warren, W. H., Jr., & Hannon, D. J. (1990). Eye movements and optical flow. *Journal of the Optical Society of America A, 7,* 160–169.

Waxman, A. M., & Wohn, K. (1987). Contour evolution, neighborhood deformation and image flow: textured surfaces in motion. In W. Richards & S. Ullman (Eds.). *Image Understanding 1985-1986* (pp. 72–98). Norwood, NJ: Ablex Publ.

Weng, J., Huang, T. S., & Ahuja, N. (1989). Motion and structure from two perspective views: algorithms, error analysis, and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11,* 451–476.

Whitney, H. (1955). On singularities of mappings of euclidean spaces. I. mappings of the plane into the plane. *Annals of Mathematics, 62,* 374–410.

Wolff, L. B., & Boult, T. E. (1989). Using line correspondence stereo to measure surface orientation. In *Proceedings 11th International Joint Conference on Artificial Intelligence* (pp. 1655–1660). San Mateo, CA: Kaufmann.

Yonas, A., Craton, L. G., & Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. *Perception & Psychophysics, 41,* 53–59.

Appendix A

Selected algorithms and proofs

A.1 Coordinate systems

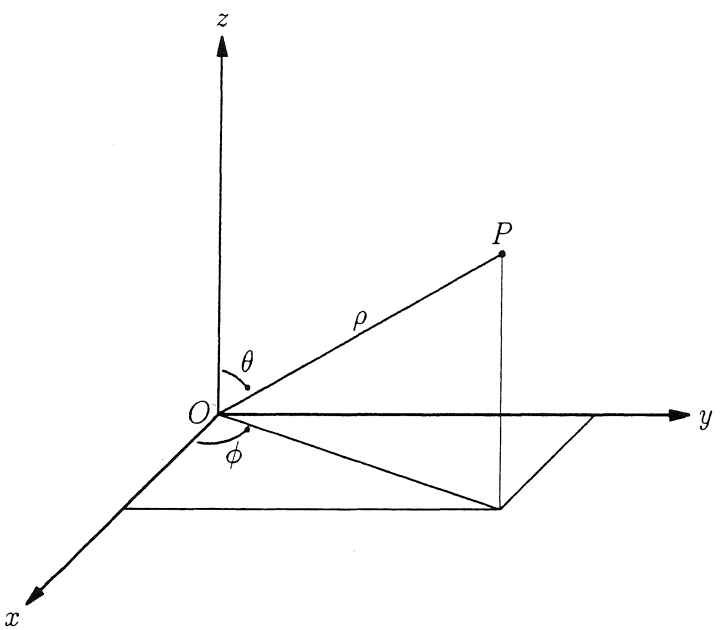


Figure A.1. The cartesian (x, y, z) and spherical (ρ, θ, ϕ) coordinates of a point P are:

$$x = \rho \sin\theta \cos\phi \qquad \rho = (x^2 + y^2 + z^2)^{1/2}$$
$$y = \rho \sin\theta \sin\phi \qquad \theta = \tan^{-1}[(x^2 + y^2)^{1/2}/z]$$
$$z = \rho \cos\theta \qquad \phi = \tan^{-1}(y/x)$$

A.2 Rotating ellipsoids

Let E be an arbitrary ellipsoid

$$E: \quad a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz = 1. \tag{A.1}$$

We will rotate E about an arbitrary unit vector $\mathbf{u} = (u_1, u_2, u_3)$ by θ degrees. The corresponding rotation matrix R, with $c_\theta = \cos\theta$ and $s_\theta = \sin\theta$, is (Faux & Pratt, 1979):

$$R = \begin{pmatrix} u_1^2 + c_\theta(1-u_1^2) & u_1u_2(1-c_\theta) - u_3s_\theta & u_3u_1(1-c_\theta) + u_2s_\theta \\ u_1u_2(1-c_\theta) + u_3s_\theta & u_2^2 + c_\theta(1-u_2^2) & u_2u_3(1-c_\theta) - u_1s_\theta \\ u_3u_1(1-c_\theta) - u_2s_\theta & u_2u_3(1-c_\theta) + u_1s_\theta & u_3^2 + c_\theta(1-u_3^2) \end{pmatrix}. \tag{A.2}$$

Under R, a point $\mathbf{x} = (x, y, z)$ will be rotated to $\mathbf{x}' = (x', y', z') = R\mathbf{x}$. Because R is orthogonal, $R^{-1} = R^T$. Substituting $\mathbf{x} = R^{-1}\mathbf{x}' = R^T\mathbf{x}'$ into Equation A.1, and collecting terms we obtain E':

$$E': \quad a'_{11}x'^2 + a'_{22}y'^2 + a'_{33}z'^2 + 2a'_{12}x'y' + 2a'_{13}x'z' + 2a'_{23}y'z' = 1, \tag{A.3}$$

with

$$\begin{aligned}
a'_{11} &= a_{11}r_{11}^2 + a_{22}r_{12}^2 + a_{33}r_{13}^2 + 2a_{12}r_{11}r_{12} + 2a_{13}r_{11}r_{13} + 2a_{23}r_{12}r_{13} \\
a'_{22} &= a_{11}r_{21}^2 + a_{22}r_{22}^2 + a_{33}r_{23}^2 + 2a_{12}r_{21}r_{22} + 2a_{13}r_{21}r_{23} + 2a_{23}r_{22}r_{23} \\
a'_{33} &= a_{11}r_{31}^2 + a_{22}r_{32}^2 + a_{33}r_{33}^2 + 2a_{12}r_{31}r_{32} + 2a_{13}r_{31}r_{33} + 2a_{23}r_{32}r_{33} \\
a'_{12} &= a_{11}r_{11}r_{21} + a_{22}r_{12}r_{22} + a_{33}r_{13}r_{23} \\
&\quad + a_{12}(r_{11}r_{22} + r_{12}r_{21}) + a_{13}(r_{11}r_{23} + r_{21}r_{13}) + a_{23}(r_{12}r_{23} + r_{22}r_{13}) \\
a'_{13} &= a_{11}r_{11}r_{31} + a_{22}r_{12}r_{32} + a_{33}r_{13}r_{33} \\
&\quad + a_{12}(r_{11}r_{32} + r_{31}r_{12}) + a_{13}(r_{11}r_{33} + r_{31}r_{13}) + a_{23}(r_{12}r_{33} + r_{32}r_{13}) \\
a'_{23} &= a_{11}r_{21}r_{31} + a_{22}r_{22}r_{32} + a_{33}r_{23}r_{33} \\
&\quad + a_{12}(r_{21}r_{32} + r_{31}r_{22}) + a_{13}(r_{21}r_{33} + r_{31}r_{23}) + a_{23}(r_{22}r_{33} + r_{32}r_{23})
\end{aligned}$$

where r_{ij} denotes the element in the i-th row and j-th column of R.

A.3 Translating ellipsoids

We will translate the ellipsoid E by $\delta\mathbf{x} = (\delta x, \delta y, \delta z)$. Substitute $\mathbf{x} = \mathbf{x}' + \delta\mathbf{x}$ into Equation A.1, and collect terms:

$$\begin{aligned}
E': \quad & a'_{11}x'^2 + a'_{22}y'^2 + a'_{33}z'^2 + 2a'_{12}x'y' + 2a'_{13}x'z' + 2a'_{23}y'z' \\
& + 2a'_1x + 2a'_2y + 2a'_3z + a'_0 = 1,
\end{aligned} \tag{A.4}$$

with

$$
\begin{aligned}
a'_{11} &= a_{11}, \quad a'_{22} = a_{22}, \quad a'_{33} = a_{33}, \\
a'_{12} &= a_{12}, \quad a'_{13} = a_{11}, \quad a'_{23} = a_{11}, \\
a'_1 &= a_{11}\delta x + a_{12}\delta y + a_{13}\delta z, \\
a'_2 &= a_{12}\delta x + a_{22}\delta y + a_{23}\delta z, \\
a'_3 &= a_{13}\delta x + a_{23}\delta y + a_{33}\delta z, \\
a'_0 &= a_{11}\delta^2 x + a_{22}\delta^2 y + a_{33}\delta^2 z + 2a_{12}\delta x\delta y + 2a_{13}\delta x\delta z + 2a_{23}\delta y\delta z.
\end{aligned}
$$

A.4 Rims and contours of ellipsoids

Let $\mathbf{N}(\mathbf{x})$ be the unit normal to the ellipsoid defined by Equation A.4 and drop the latter's primes for clarity. $\mathbf{N} = \mathbf{n}/|\mathbf{n}|$, where

$$
\begin{aligned}
\mathbf{n} = (&a_{11}x + a_{12}y + a_{13}z + a_1, \\
&a_{12}x + a_{22}y + a_{23}z + a_2, \\
&a_{13}x + a_{23}y + a_{33}z + a_3).
\end{aligned} \tag{A.5}
$$

Let the vantage point be at $(0, 0, z_V)$; the viewing direction associated with the point (x, y, z) is then $\mathbf{V} = (x, y, z - z_V)/(x^2 + y^2 + (z - z_V)^2)^{1/2}$. The rim of the ellipsoid is defined by $(\mathbf{N} \cdot \mathbf{V}) = 0$, which allows us to solve for z:

$$
\begin{aligned}
(\mathbf{N} \cdot \mathbf{V}) &= a_{11}x^2 + a_{12}xy + a_{13}xz + a_1xa_{12}xy + a_{22}y^2 + a_{23}yz + a_2y \\
&\quad + a_{13}x(z - z_V) + a_{23}y(z - z_V) + a_{33}z(z - z_V) + a_3(z - z_V) \\
&= a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz \\
&\quad + a_1x + a_2y + a_3z - z_V(a_{13}x + a_{23}y + a_{33}z + a_3) \\
&= 1 - a_1x - a_2y - a_3z - z_V(a_{13}x + a_{23}y + a_{33}z + a_3) \\
&= 0.
\end{aligned} \tag{A.6}
$$

Thus,

$$
\begin{aligned}
z &= \frac{1 - (a_1 + z_V a_{13})x - (a_2 + z_V a_{23})y - z_V a_3}{a_3 + z_V a_{33}} \\
&= \frac{1 - a'_{13}x - a'_{12}y - z_V a_3}{a'_{33}}
\end{aligned} \tag{A.7}
$$

with $a'_{13} = a_1 + z_V a_{13}$, $a'_{12} = a_2 + z_V a_{23}$, and $a'_{33} = a_3 + z_V a_{33}$.

Equations A.4 and A.7 together define the rim of the ellipsoid E as seen from the vantage point $(0, 0, z_V)$. Note that Equation A.7 defines a plane. In other words, the rim of an ellipsoid is planar under perspective projection, and, a fortiori, under orthographic projection.

To compute the projection of the rim, i.e., the occluding contours $\mathcal{C}$, substitute the expression for z defined by Equation A.7 into Equation A.4:

$$\begin{aligned}\mathcal{C}: \quad & a_{11}x^2 + a_{22}y^2 + a_{33}\frac{(1 - a'_{13}x - a'_{23}y - z_V a_3)^2}{(a'_{33})^2} + 2a_{12}xy \\ & + 2a_{13}x\frac{(1 - a'_{13})x - a'_{23}y - z_V a_3)}{a'_{33}} \\ & + 2a_{23}y\frac{(1 - a'_{13})x - a'_{23}y - z_V a_3)}{a'_{33}} \\ & + 2a_1 x + 2a_2 y + 2a_3\frac{(1 - a'_{13})x - a'_{23}y - z_V a_3)}{a'_{33}} + a_0 = 1.\end{aligned} \tag{A.8}$$

Collecting terms,

$$\begin{aligned}\mathcal{C}: \quad & (a_{11} + \frac{(a'_{13})^2}{(a'_{33})^2} - 2\frac{a'_{13}}{a'_{33}}a_{13})x^2 \\ & + (a_{22} + \frac{(a'_{23})^2}{(a'_{33})^2}a_{33} - 2\frac{a'_{23}}{a'_{33}}a_{23})y^2 \\ & + 2(\frac{a_{33}a'_{13}a'_{23}}{(a'_{33})^2} + a_{12} - \frac{a_{13}a'_{23}}{a'_{33}} - \frac{a_{23}a'_{13}}{a'_{33}})xy \\ & + 2(-a_{33}\frac{(1 - z_V a_3)a'_{13}}{(a'_{33})^2} + a_{13}\frac{(1 - z_V a_3)}{a'_{33}} + a_1 - a_3\frac{a'_{13}}{a'_{33}})x \\ & + 2(-a_{33}\frac{(1 - z_V a_3)a'_{23}}{(a'_{33})^2} + a_{23}\frac{(1 - z_V a_3)}{a'_{33}} + a_2 - a_3\frac{a'_{23}}{a'_{33}})y \\ & + a_{33}\frac{(1 - z_V a_3)^2}{(a'_{33})^2} + 2a_3\frac{(1 - z_V a_3)}{a'_{33}} + a_0 = 1.\end{aligned} \tag{A.9}$$

If the ellipsoid is only being rotated, its contours are given by:

$$\mathcal{C}: \quad (a_{11} - \frac{a_{13}^2}{a_{33}})x^2 + (a_{22} - \frac{a_{23}^2}{a_{33}})y^2 + 2(a_{12} - \frac{a_{13}a_{23}}{a_{33}})xy + \frac{1}{a_{33}z_V^2} = 1. \tag{A.10}$$

Note that the contours under perspective projection (z_V is finite) are simply scaled versions of the contours under orthographic projection ($z_V = \infty$), the scaling factor in x and y being $[(z_V^2 a_{33} - 1)/z_V^2 a_{33}]^{1/2}$.

A.5 Envelope of ellipse rotating in the plane

Let Φ represent an ellipse E in the XY-plane, rotating about the Z-axis (with $c_\theta = \cos\theta$ and $s_\theta = \sin\theta$):

$$\Phi: \quad a_{11}(xc_\theta + ys_\theta)^2 + a_{22}(yc_\theta - xs_\theta)^2 - 1 = 0. \tag{A.11}$$

The envelope of E is given by $\Phi = \partial\Phi/\partial\theta = 0$.

$$\begin{aligned}\frac{\partial\Phi}{\partial\theta} &= 2a_{11}(xc_\theta + ys_\theta)(-xs_\theta + yc_\theta) + 2a_{22}(yc_\theta - xs_\theta)(-ys_\theta - xc_\theta) \\ &= (xc_\theta + ys_\theta)(yc_\theta - xs_\theta)(a_{11} - a_{22}) = 0.\end{aligned} \tag{A.12}$$

If $a_{11} = a_{22}$, i.e., E is a circle, all points of E lie on the envelope. If $xc_\theta + ys_\theta = 0$, the envelope is given by $a_{22}(yc_\theta - xs_\theta)^2 = 1$ and $xc_\theta + ys_\theta = 0$. Eliminating x yields $y = \pm\sqrt{a_{22}}\,c_\theta$, and eliminating y yields $x = \pm\sqrt{a_{22}}\,s_\theta$. The envelope is thus a circle with radius $\sqrt{a_{22}}$. The remaining possibility, $yc_\theta - xs_\theta = 0$, leads to a circle with radius $\sqrt{a_{11}}$.

A.6 Computing the differential invariants of the optic flow field

Assume we are looking at a densely textured object and consider two successive views of three neighboring points that are not collinear: $(O, P_1, P_2) \rightarrow (O', P_1', P_2')$. Translate the second view such that $O' = O$; let $Q_1 = P_1' - (O - O')$ and $Q_2 = P_2' - (O - O')$. Thus, $\triangle(O, P_1, P_2) \rightarrow \triangle(O, Q_1, Q_2)$. This change in shape can be described by the 2-by-2 matrix M defined by $MP_1 = Q_1$ and $MP_2 = Q_2$ in some orthogonal coordinate system x-y centered at O. Thus we have four linear equations in four unknowns:

$$\begin{aligned} m_{00}P_{1x} + m_{01}P_{1y} &= Q_{1x} \\ m_{10}P_{1x} + m_{11}P_{1y} &= Q_{1y} \\ m_{00}P_{2x} + m_{01}P_{2y} &= Q_{2x} \\ m_{10}P_{2x} + m_{11}P_{2y} &= Q_{2y} \end{aligned} \tag{A.17}$$

where m_{ij} is the element in the i-th row and j-th column of M.

The three first-order differential invariants divergence (div), curl, and deformation (def) describe the change in shape in a coordinate system free manner. The differential invariants are:

$$\begin{aligned} \text{div} &= m_{00} + m_{11}, \\ \text{curl} &= m_{10} - m_{01}, \\ \text{def} &= \sqrt{(m_{00} - m_{11})^2 + (m_{01} + m_{10})^2}, \\ \delta &= \frac{1}{2}\text{atan}\left(\frac{m_{01} + m_{10}}{m_{00} - m_{11}}\right), \end{aligned} \tag{A.19}$$

where δ is the angle between the local x-axis and the axis of contraction.

The optic flow associated with the image of a surface patch is determined by the movement of the observer relative to the patch and the orientation of the surface patch. Let the observer translate by $\mathbf{T}$ and rotate by $\mathbf{R}$ about an axis through the vantage point, and let the surface patch be in the direction $\mathbf{r}$ at a distance ρ from the observer. Decompose $\mathbf{T}$ and $\mathbf{R}$ into components along $\mathbf{r}$ (yielding T_r and R_r) and perpendicular to $\mathbf{r}$ (yielding $\mathbf{T}_t = \mathbf{T} - T_r\mathbf{r}$ and $\mathbf{R}_t = \mathbf{R} - R_r\mathbf{r}$). Define $\mathbf{A}_t = \mathbf{T}_t/\rho$ and $A_r = T_r/\rho$. If we describe the surface orientation by $\mathbf{F} = \text{grad}\log(\rho_0/\rho)$, then

$$\begin{aligned} \text{div} &= -\mathbf{F}\cdot\mathbf{A}_t + 2A_r, \\ \text{curl} &= -\mathbf{F}\times\mathbf{A}_t - 2R_r, \\ \text{def} &= FA_t, \end{aligned} \tag{A.20}$$

where $F = |\mathbf{F}|$ and $A_t = |\mathbf{A}_t|$; the axis of contraction of the deformation bisects $\mathbf{F}$ and $\mathbf{A}_t$.

For more details regarding differential invariants see Koenderink (1985, 1986) and Koenderink & van Doorn (1975, 1976c, 1986).

A.7 Computing conjugation relationships from the slant field

The purpose of this section is to prove that for any two neighboring points p and q in the image, the projection of the direction conjugated to $\overrightarrow{PQ}$, with $P = \pi^{-1}(p)$ and $Q = \pi^{-1}(q)$, is perpendicular to $d\mathbf{F} = \mathbf{F}_q - \mathbf{F}_p$, the difference in slant at p and q. As a short hand, I will say that $\overrightarrow{pq}$ is conjugated to $d\mathbf{F}$. The approach is to start with a Monge-like patch for which the assertion holds, and to show that arbitrary rotations of the patch do not affect the assertion.

The tangent plane at P is T_P and at Q it is T_Q; the corresponding normals are $\mathbf{N}_P$ and $\mathbf{N}_Q$. Let $\mathbf{E}$ be the edge along which T_P and T_Q intersect. $\mathbf{E}$ and $\overrightarrow{PQ}$ are conjugated directions.

Given $\overrightarrow{pq}$, we want to find the direction conjugated to it, that is, the direction $\mathbf{e} = \pi(\mathbf{E})$, the projection of $\mathbf{E}$. Choose a coordinate frame centered at the point midway between P and Q such that its x-axis is parallel to $d\mathbf{F}$, its y-axis is parallel to $\mathbf{e}$ and its z-axis parallel to $\mathbf{N}_P + \mathbf{N}_Q$ (Figure A.2a). Expressing the two surface normals in this coordinate frame: $\mathbf{N}_P = (-N_x, 0, N_z)$ and $\mathbf{N}_Q = (N_x, 0, N_z)$. As the slant at a point is $(\delta z/\delta x, \delta z/\delta y)$, we have $\mathbf{F}_p = (-N_x/N_z, 0)$, $\mathbf{F}_q = (N_x/N_z, 0)$, and $d\mathbf{F} = \mathbf{F}_q - \mathbf{F}_p = (2N_x/N_z)$. The difference in normal is $d\mathbf{N} = \mathbf{N}_Q - \mathbf{N}_P = (2N_x, 0, 0)$. Note that $\mathbf{e}\cdot d\mathbf{F} = (0,1)\cdot(2N_x/N_Z, 0) = 0$; similarly, $\mathbf{e}\cdot\pi(d\mathbf{N}) = (0,1)\cdot(2N_x, 0) = 0$. Thus, the difference in slant as well as the difference in the projection of the surface normals is perpendicular to $\mathbf{e}$.

Obviously, rotations about the z-axis result in a mere relabeling of image directions and hence do not affect the relationship between $\overrightarrow{pq}$ and $d\mathbf{F}$ (Figure A.2b). In the following I assume that the x- and y-axes have been renamed to be once again parallel to $d\mathbf{F}$ and $\mathbf{e}$, respectively.

Rotate the tangent planes α degrees about an axis perpendicular to $\mathbf{E}$ (Figure A.2c). Then $\mathbf{N}'_P = (-N_x, N_z s_\alpha, N_z c_\alpha)$, $\mathbf{N}'_Q = (N_x, N_z s_\alpha, N_z c_\alpha)$, and $\mathbf{E}' = (0, c_\alpha, -s_\alpha)$. The slants are $\mathbf{F}'_p = (-N_x/N_z c_\alpha, t_\alpha)$, and $\mathbf{F}'_q = (N_x/N_z c_\alpha, t_\alpha)$, where $t_\alpha = \tan\alpha$. Thus, $\mathbf{e}'\cdot d\mathbf{F}' = (0, c_\alpha)\cdot(2N_x/N_z c_\alpha, 0) = 0$, and $\mathbf{e}'\cdot\pi(d\mathbf{N}') = (0, c_\alpha)\cdot(2N_x, 0) = 0$. In other words, the

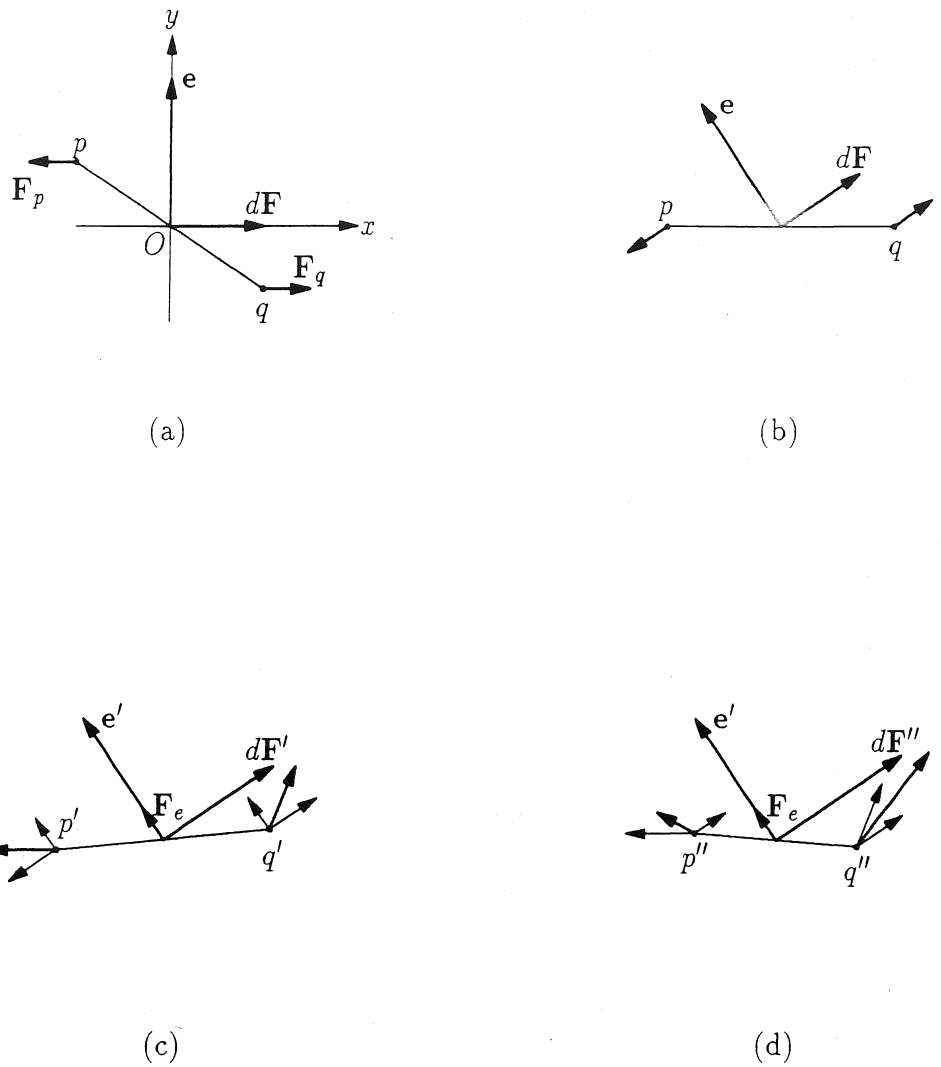


Figure A.2. The direction conjugated to $\overrightarrow{pq}$ is perpendicular to $d\mathbf{F}$. (a) A Monge-like patch: the x-direction is ligned up with $d\mathbf{F}$, the y-direction with $\mathbf{e}$, the z-direction is parallel to $\mathbf{N}_P + \mathbf{N}_Q$. (b) After rotating the patch 33° about the z-axis. (c) After rotating the patch α degrees about the x-axis, the slant in the direction of $\mathbf{e}$ is $\mathbf{F}_e$. The slant at p and q increases by a factor $1/\cos(\alpha)$, as well as by the addition of $\mathbf{F}_e$. (d) Rotation about $\mathbf{E}'$ changes the slant at p and q only in the direction of $d\mathbf{F}'$, so that $\mathbf{e}'$ remains perpendicular to $d\mathbf{F}' = d\mathbf{F}''$.

difference in slant as well as the difference in the projection of the surface normals remains perpendicular to $\mathbf{e}'$.

Now rotate the tangent planes β degrees about $\mathbf{E}'$ (Figure A.2d). The corresponding rotation matrix $R_{E'}$ is:

$$R_{E'} = \begin{pmatrix} c_\beta & s_\alpha s_\beta & c_\alpha s_\beta \\ -s_\alpha s_\beta & c_\alpha^2 + c_\beta s_\alpha^2 & -c_\alpha s_\alpha (1 - c_\beta) \\ -c_\alpha s_\beta & -c_\alpha s_\alpha (1 - c_\beta) & s_\alpha^2 + c_\beta c_\alpha^2 \end{pmatrix}. \tag{A.21}$$

The new normals are $\mathbf{N}''_P = R_{E'}\mathbf{N}'_P$ and that $\mathbf{N}''_Q = R_{E'}\mathbf{N}'_Q$:

$$\mathbf{N}''_P = \begin{pmatrix} -N_x c_\beta + N_z s_\beta \\ s_\alpha(N_x s_\beta + N_z c_\beta) \\ c_\alpha(N_x s_\beta + N_z c_\beta) \end{pmatrix}, \quad \mathbf{N}''_Q = \begin{pmatrix} N_x c_\beta + N_z s_\beta \\ s_\alpha(-N_x s_\beta + N_z c_\beta) \\ c_\alpha(-N_x s_\beta + N_z c_\beta) \end{pmatrix}. \tag{A.22}$$

The new slants are:

$$\mathbf{F}''_p = \begin{pmatrix} \frac{-N_x c_\beta + N_z s_\beta}{c_\alpha(N_x s_\beta + N_z c_\beta)} \\ t_\alpha \end{pmatrix}, \quad \mathbf{F}''_q = \begin{pmatrix} \frac{N_x c_\beta + N_z s_\beta}{c_\alpha(-N_x s_\beta + N_z c_\beta)} \\ t_\alpha \end{pmatrix}. \tag{A.23}$$

The difference in normals and slants is:

$$d\mathbf{N}'' = \begin{pmatrix} 2N_x c_\beta \\ -2N_x s_\alpha s_\beta \end{pmatrix}, \quad d\mathbf{F}'' = \begin{pmatrix} \frac{2N_x N_z}{c_\alpha(N_z^2 c_\beta^2 - N_x^2 s_\beta^2)} \\ 0 \end{pmatrix}. \tag{A.24}$$

It follows that the difference in slant at P and Q is perpendicular to $\mathbf{e}' = (0, 1)$, but that the difference in their normals is not.

A.8 Velocity field induced by rotating objects

The purpose of this section is to derive expressions for the velocity field arising from a rotating rigid object under perspective projection. Let the observer be at the origin of the spherical coordinate system shown in Figure A.1; let the object be rotating about a fixed axis $\Omega = (\omega_x, \omega_y, \omega_z)$ through the point $R = (R_x, R_y, R_z)$, and let $\mathbf{r} = (x, y, z) = (\rho, \theta, \phi)$ specify a point on the object's surface. The velocity of $\mathbf{r}$ is then

$$\begin{aligned} \dot{\mathbf{r}} &= \Omega \times (\mathbf{r} - R) \\ &= \begin{vmatrix} \omega_x & \omega_y & \omega_z \\ x - R_x & y - R_y & z - R_z \\ e_1 & e_2 & e_3 \end{vmatrix} \\ &= \big(\omega_y(z - R_z) - \omega_z(y - R_y)\big)\mathbf{e}_1 \\ &\quad - \big(\omega_x(z - R_z) - \omega_z(x - R_x)\big)\mathbf{e}_2 \\ &\quad + \big(\omega_x(y - R_y) - \omega_y(x - R_x)\big)\mathbf{e}_3, \end{aligned} \tag{A.25}$$

where $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{e}_3$ are unit vectors in the direction of the x, y, and z axes, respectively.

The image velocity is then $(\dot{\theta}, \dot{\phi}) = (\dot{\mathbf{r}} \cdot \nabla\theta, \dot{\mathbf{r}} \cdot \nabla\phi)$, where

$$\nabla\theta = \rho^{-1} \begin{pmatrix} c_\theta c_\phi \\ c_\theta s_\phi \\ -s_\theta \end{pmatrix}, \qquad \text{and} \qquad \nabla\dot{\phi} = (\rho s_\theta)^{-1} \begin{pmatrix} -s_\phi \\ c_\phi \\ 0 \end{pmatrix}. \tag{A.26}$$

Thus,

$$\begin{aligned} u = \dot{\phi} = {} & \omega_x (R'_z c_\phi - c_\theta c_\phi)/s_\theta \\ & + \omega_y (R'_z s_\phi - c_\theta s_\phi)/s_\theta \\ & + \omega_z \left(1 - (R'_x c_\phi + R'_y s_\phi)/s_\theta\right), \end{aligned} \tag{A.27}$$

and

$$\begin{aligned} v = \dot{\theta} = {} & \omega_x (R'_y s_\theta + R'_z c_\theta s_\phi - s_\phi) \\ & + \omega_y (c_\phi - R'_x s_\theta - R'_z c_\theta c_\phi) \\ & + \omega_z c_\theta (R'_y c_\phi - R'_x s_\phi), \end{aligned} \tag{A.28}$$

where $R'_x = R_x/\rho$, etc.

The corresponding equations for an observer translating with constant linear velocity $\mathbf{T} = (v_x, v_y, v_z)$ while rotating about an axis ϖ through the origin are (Rieger, 1983):

$$\begin{aligned} u = \dot{\phi} &= (v_x s_\phi - v_y c_\phi)/s_\theta \rho + (\varpi_x c_\phi + \varpi_y s_\phi) c_\theta / s_\theta - \varpi_z, \\ v = \dot{\theta} &= (-v_x c_\theta c_\phi - v_y c_\theta s_\phi + v_z s_\theta)/\rho + \varpi_x s_\phi - \varpi_y c_\phi. \end{aligned} \tag{A.29}$$

Thus, the velocity field of a rotating object plus stationary observer is equivalent to that of a stationary object and an observer who rotates about $\varpi = -\Omega$ and translates by $\mathbf{T} = (R_z\omega_y - R_y\omega_z, R_x\omega_z - R_z\omega_x, R_y\omega_x - R_x\omega_y)$.

Consider the special case of an object rotating about an axis which is parallel to the vertical and located straight ahead; in other words, $\Omega = (0, 0, \omega_z)$ and $R = (0, R_y, 0)$. Then,

$$\begin{aligned} u = \dot{\phi} &= -\omega_z (R_y s_\phi / s_\theta \rho - 1), \\ v = \dot{\theta} &= \omega_z R_y c_\theta c_\phi / \rho. \end{aligned} \tag{A.30}$$

The same expressions obtain for a stationary object and an observer who translates by $\mathbf{T} = (-R_y\omega_z, 0, 0)$ while rotating about $\varpi = (0, 0, -\omega_z)$.

A.9 Velocity field of a rotating plane

Consider the situation depicted in Figure A.3. An observer at O is looking at a plane $\mathcal{P}$ rotating about a vertical axis Ω and at a distance R_y straight ahead; $\mathcal{P}$ is slanted by σ. Suppose the plane is textured and induces a velocity field on the observer's optic array. This velocity field obviously depends on whether we are using orthographic or perspective

projection. But as I will show in this section the difference between the two is actually quite small.

Starting with perspective projection, the general expression for the velocity field $(u(\theta,\phi), v(\theta,\phi))$, where u is the velocity in the direction of $\mathbf{e}_\phi$ and v is in the direction of $\mathbf{e}_\theta$, was derived in section A.8, and is given in Equation A.30. All we need to do now is derive an expression for ρ. The points on $\mathcal{P}$ are given by $z = \cot\sigma(c_\omega(y - R_y) - s_\omega x)$, except when $\sigma = 0$ (ω is the amount of rotation about Ω and should not be confused with ω_z, the angular *velocity* about Ω). Substituting $x = \rho s_\theta c_\phi, y = \rho s_\theta s_\phi, z = \rho c_\theta$ and solving for ρ^{-1} yields:

$$\rho^{-1} = -\frac{\tan\sigma c_\theta - s_\theta(s_\phi c_\omega - c_\phi s_\omega)}{R_y c_\omega}. \tag{A.31}$$

In the special case $\tan\sigma = 0$, the plane $\mathcal{P}$ is defined by $y = R_y$. Solving for ρ^{-1} again yields Equation A.31 if we set $\tan\sigma = 0$. Thus, we can use Equation A.31 throughout. Substituting Equation A.31 into Equation A.30 gives:

$$\begin{aligned} u &= \omega_z[\frac{s_\phi}{c_\omega}(\tan\sigma\cot\theta - s_\phi c_\omega + c_\phi s_\omega) + 1], \\ v &= -\omega_z c_\theta c_\phi[\tan\sigma c_\theta - s_\theta(s_\phi c_\omega - c_\phi s_\omega)]/c_\omega. \end{aligned} \tag{A.32}$$

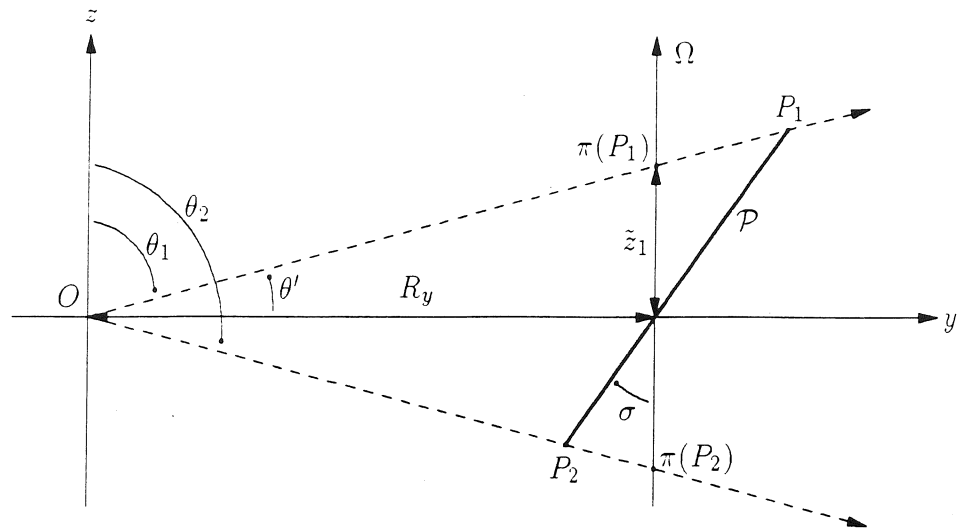
Assuming that the field of view is only a few degrees in diameter, we can use small angle approximations. Keeping in mind that ϕ and θ are both close to 90°, and letting $\phi' = 90° - \phi$ and $\theta' = 90° - \theta$, we obtain $\cos(\phi) = \cos(90° - \phi') = \sin(\phi') \approx \phi'$, and $\sin(\phi) = \sin(90° - \phi') = \cos(\phi') \approx 1$. Similarly, $\cos(\theta) \approx \theta'$, $\sin(\theta) \approx 1$, and $\cot(\theta) \approx \theta'$. Note that the ϕ'-axis is parallel to the x-axis, and that the θ'-axis is parallel to the z-axis. This results in:

$$\begin{aligned} u^1 &= \frac{\omega_z}{c_\omega}(\theta'\tan\sigma + \phi' s_\omega) \\ v^1 &= -\frac{\omega_z}{c_\omega}\theta'\phi'(\theta'\tan\sigma - c_\omega + \phi' s_\omega) \end{aligned} \tag{A.33}$$

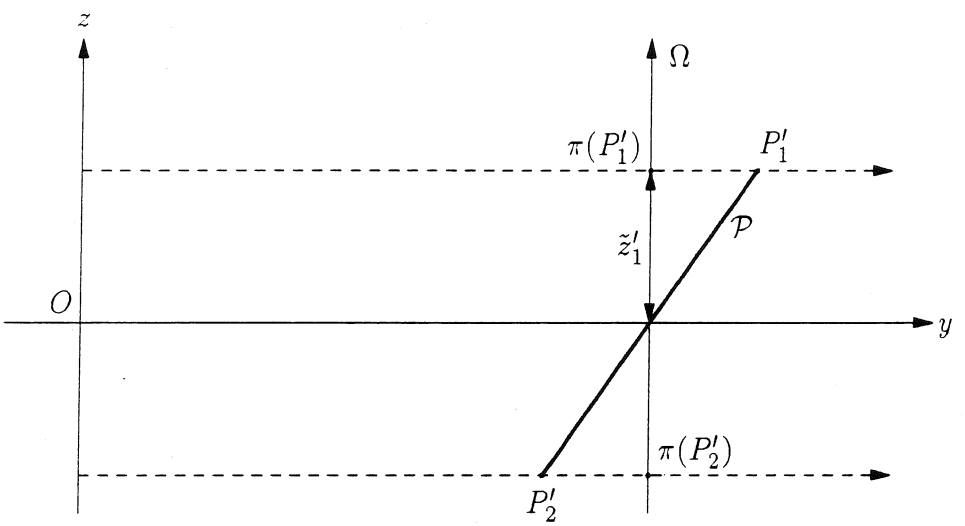
The velocity v^1 can be further simplified to $\omega_z\theta'\phi'$ for small slants ($\sigma < 40°$) and small angles ω_z ($\omega_z < 60°$).

Table A.1 illustrates that the first-order approximation of Equation A.33 is quite accurate up to a field of view of 10° in diameter; then the approximate velocity u^1 starts to deviate appreciably from the real velocity u, especially for small slants; the velocity v^1 remains within .5% from v. (There are many ways to compare two velocity fields; perhaps the best way would have been to compute the mean of the ratios u/u^1 and v/v^1; the problem, however, is that u^1 and v^1 approach zero in some locations causing the ratio to become arbitrarily large. Therefore I tabulated the ratio of the mean of $|u|$ and the mean of $|u - u^1|$ in Table A.1.)

The velocity field of Equation A.33 is very close to what we would get under orthographic projection. In that case, $\dot{\mathbf{r}} = \omega_z(R_y - y, x, 0)$, as before; $\nabla\theta = (0, 0, -1/\rho)$ and $\nabla\phi = (-1/\rho, 0, 0)$. Thus, $u^o = \dot{\mathbf{r}}\cdot\nabla\phi = -\omega_z(R_y - y)/\rho = -\omega_z(R_y/\rho - 1)$, and $v^o = \dot{\mathbf{r}}\cdot\nabla\theta = 0$. $\rho^{-1} = -(\theta'\tan\sigma - c_\omega + \phi' s_\omega)/R_y c_\omega$, so that $u^o = \omega_z(\theta'\tan\sigma + \phi' s_\omega)/c_\omega$.



(a)



(b)

Figure A.3 Side view of a plane $\mathcal{P}$ rotating about the axis Ω going through $(0, R_y, 0)$. The projection plane is parallel to the $x - z$ plane and also passes through $(0, R_y, 0)$. The plane is slanted by σ degrees with respect to the $x - z$ plane. (a) Perspective projection. The viewing direction $r = (\theta, \phi)$ of P_1 is $r_1 = (\theta_1, 90°)$; similarly, $r_2 = (\theta_2, 90°)$. $\theta_1 = 90° - \theta'$ and $\theta_2 = 90° + \theta'$. (b) Orthographic projection. Image velocities at $\pi(P_1)$ and $\pi(P_1')$ are equal and opposite to the velocities at $\pi(P_2)$ and $\pi(P_2')$, which are also equal.

Table A.2 illustrates the difference between the velocity field of a slanted plane rotating about an axis parallel to the vertical, as shown in Figure A.3, under perspective and orthographic projection. The plane rotates at $15°/s$; the values in the table are in minutes of visual angle per second. Note that the difference is of the same order of magnitude for u and v, i.e., for the velocity in the horizontal and vertical directions, respectively. Furthermore, the difference increases with approximately the square of the distance from the center of the visual field. It is remarkable that the difference depends very little on the amount of rotation of the plane (ω).

Figure A.3 illustrates the case in which $\omega = 0$ and $\phi' = 0$. Consider the points P_1 and P_2 in Figure A.3a. Their respective viewing directions are $(\theta', 0)$ and $(-\theta', 0)$, so that, according to Equation A.33, $u(\pi(P_1)) = -u(\pi(P_2))$. It is interesting that the effects of distance from the observer and distance from the axis of rotation compensate each other, so that P_1 which is farther from Ω (and hence has a larger 3D velocity than P_2), produces the same image velocity as P_2, which is closer to the observer (and hence has a larger image velocity given the same 3D velocity).

Now imagine that you are generating an image sequence on a CRT screen which is parallel to the $x - z$ plane and contains Ω. Under perspective projection, $\pi(P_1) = (0, \tilde{z}_1)$, in (x, z) coordinates, and $\pi(P_2) = (0, \tilde{z}_2) = (0, -\tilde{z}_1)$. Thus the same image sequence would be obtained if we had used orthographic projection and a slightly different plane, as shown in Figure A.3b ($\tilde{z}'_1 = \tilde{z}_1$). Note that the slant of both planes is the same. Table 5.1 shows the differential invariants divergence, curl, and deformation for this plane under perspective and orthographic projection (look under $R_d = 10$ and $P_d = 10$); they are identical except that there is some variation in divergence as a function of θ (not shown in Table; see Chapter 5 for explanation). In Chapter 5, I also discuss experiments in which subjects have to judge the slant of a rotating plane; I found no appreciable difference between perspective and orthographic projection.

Although v is much smaller than u, it is not entirely negligible if the field of view is larger than just a few degrees. For example, it $\omega_z = 15°/s$, and $\theta' = \phi' = 5°$, then $v = -\omega_z \theta' \phi' = -6.9'/s$ (v is negative because $\mathbf{e}_\theta = (c_\theta c_\phi, c_\theta s_\phi, -s_t) \approx= (\theta' \phi', \theta', -1)$, i.e., downwards in the image). This will cause horizontal lines to change their orientation; non-horizontal lines change their orientation even if $v = 0$. Perhaps a better way to appreciate this is to follow a fixed point as it rotates about Ω and to compute its θ' along the way. Let $P = (0, R_y + z \tan\sigma, z)$ be such a point; as it rotates about Ω its coordinates become $P' = (z \tan\sigma s_\omega, R_y + z \tan\sigma c_\omega, z)$, and $\theta' = 90° - \tan^{-1}[(x^2 + y^2)^{1/2}/z] = 90° - \tan^{-1}[z^2 \tan\sigma^2 + R_y^2 + 2R_y z \tan\sigma c_\omega]^{1/2}/z$. If $R_y = 10, z = 1$, and $\sigma = 30°$, then θ' increases from 5.4 to 6.1° as ω goes from 0 to 180°. The corresponding numbers for $\sigma = 50°$ are 5.1 and 6.5°; and for $\sigma = 70°$, 4.9 and 7.9°.

A.10 Deformation of the optic flow field of a cylinder translating in depth

Let $\mathcal{C}$ be a cylinder with radius r_c, whose length axis is parallel to the z-axis and goes through $(0, R_y, 0)$. We will only consider the deformation along $z = 0$, or, equivalently, $\theta = 90°$ (Figure A.1). The distance function ρ for this part of the cylinder is

$$\rho = R_y s_\phi - \sqrt{r_c^2 - c_\phi^2 R_y^2}. \qquad (A.34)$$

Table A.1 A comparison between the real image velocity field induced by a rotating plane and its first-order approximation given by Equation A.33. I computed the mean of the absolute value of the real velocity field, ($\overline{|u|}$ and $\overline{|v|}$), and the mean of the absolute value of the difference between the real velocity and its first-order approximation ($\overline{|u-u^1|}$ and $\overline{|v-v^1|}$), along 4 different circles in the image (radii .5°, 1°, 2°, and 5°), all centered at the direction straight ahead. The values in the table give the ratio of these means. The comparison was made for 5 different slant angles (σ) and after 5 different amounts of rotation (ω).

		$\overline{\lvert u-u^1\rvert}/\overline{\lvert u\rvert}$				$\overline{\lvert v-v^1\rvert}/\overline{\lvert v\rvert}$			
σ	ω	.5°	1.°	2.°	5.°	.5°	1.°	2.°	5.°
0°	0°	1.000	1.000	1.000	1.000	0.000	0.000	0.001	0.005
	20°	0.019	0.038	0.075	0.189	0.000	0.000	0.001	0.005
	40°	0.008	0.016	0.033	0.082	0.000	0.000	0.001	0.005
	60°	0.004	0.008	0.016	0.040	0.000	0.000	0.001	0.005
	80°	0.001	0.002	0.005	0.012	0.000	0.000	0.001	0.005
20°	0°	0.019	0.038	0.075	0.183	0.000	0.000	0.001	0.005
	20°	0.013	0.026	0.052	0.128	0.000	0.000	0.001	0.005
	40°	0.007	0.014	0.028	0.071	0.000	0.000	0.001	0.005
	60°	0.004	0.007	0.015	0.037	0.000	0.000	0.001	0.005
	80°	0.001	0.002	0.005	0.011	0.000	0.000	0.001	0.005
40°	0°	0.008	0.016	0.033	0.081	0.000	0.000	0.001	0.005
	20°	0.007	0.014	0.028	0.071	0.000	0.000	0.001	0.005
	40°	0.005	0.010	0.020	0.050	0.000	0.000	0.001	0.005
	60°	0.003	0.006	0.011	0.029	0.000	0.000	0.001	0.005
	80°	0.001	0.002	0.004	0.010	0.000	0.000	0.001	0.005
60°	0°	0.004	0.008	0.016	0.040	0.000	0.000	0.001	0.005
	20°	0.004	0.007	0.015	0.037	0.000	0.000	0.001	0.005
	40°	0.003	0.006	0.011	0.029	0.000	0.000	0.001	0.005
	60°	0.002	0.004	0.007	0.018	0.000	0.000	0.001	0.005
	80°	0.001	0.001	0.002	0.007	0.000	0.000	0.001	0.005
80°	0°	0.001	0.002	0.005	0.013	0.000	0.000	0.001	0.005
	20°	0.001	0.002	0.005	0.012	0.000	0.000	0.001	0.005
	40°	0.001	0.002	0.004	0.010	0.000	0.000	0.001	0.005
	60°	0.001	0.001	0.002	0.007	0.000	0.000	0.001	0.005
	80°	0.000	0.000	0.001	0.003	0.000	0.000	0.001	0.003

Table A.2 A comparison between the image velocity field induced by a slanted plane rotating about a vertical axis, under perspective and orthographic projection. I computed the mean of the absolute value of the difference in velocity under perspective projection and orthographic projection, ($\overline{|u^p - u^o|}$ and $\overline{|v^p - v^o|}$), along 4 different circles in the image (radii .5°, 1°, 2°, and 5°), all centered at the direction straight ahead. The values in the table in minutes/s; the plane rotates at 15°/s. The comparison was made for 5 different slant angles (σ) and after 5 different amounts of rotation (ω).

		$\overline{\lvert u^p - u^o \rvert}$				$\overline{\lvert v^p - v^o \rvert}$			
σ	ω	.5°	1.°	2.°	5.°	.5°	1.°	2.°	5.°
0°	0°	0.034	0.137	0.548	3.420	0.022	0.087	0.348	2.165
	20°	0.034	0.137	0.548	3.420	0.022	0.087	0.348	2.165
	40°	0.034	0.137	0.548	3.420	0.022	0.087	0.348	2.165
	60°	0.034	0.137	0.548	3.420	0.022	0.087	0.348	2.165
	80°	0.034	0.137	0.548	3.420	0.022	0.087	0.348	2.165
20°	0°	0.034	0.137	0.548	3.424	0.022	0.087	0.348	2.165
	20°	0.034	0.137	0.548	3.424	0.022	0.087	0.348	2.165
	40°	0.034	0.137	0.548	3.425	0.022	0.087	0.348	2.165
	60°	0.034	0.137	0.548	3.429	0.022	0.087	0.348	2.165
	80°	0.034	0.137	0.550	3.462	0.022	0.087	0.348	2.165
40°	0°	0.034	0.137	0.549	3.431	0.022	0.087	0.348	2.165
	20°	0.034	0.137	0.549	3.432	0.022	0.087	0.348	2.165
	40°	0.034	0.137	0.549	3.436	0.022	0.087	0.348	2.165
	60°	0.034	0.137	0.549	3.451	0.022	0.087	0.348	2.165
	80°	0.034	0.138	0.554	3.552	0.022	0.087	0.348	2.165
60°	0°	0.034	0.137	0.549	3.452	0.022	0.087	0.348	2.165
	20°	0.034	0.137	0.550	3.455	0.022	0.087	0.348	2.165
	40°	0.034	0.137	0.550	3.466	0.022	0.087	0.348	2.165
	60°	0.034	0.137	0.552	3.503	0.022	0.087	0.348	2.165
	80°	0.034	0.139	0.564	3.763	0.022	0.087	0.348	2.165
80°	0°	0.034	0.138	0.555	3.582	0.022	0.087	0.348	2.165
	20°	0.034	0.138	0.556	3.595	0.022	0.087	0.348	2.165
	40°	0.034	0.138	0.559	3.653	0.022	0.087	0.348	2.165
	60°	0.034	0.139	0.567	3.822	0.022	0.087	0.348	2.165
	80°	0.035	0.145	0.625	4.872	0.022	0.087	0.356	4.268

Let the cylinder move by $\mathbf{T} = (0, v_y, 0)$. The resulting velocity field can be obtained by substituting $\mathbf{T}' = (0, -v_y, 0)$, $\Omega = 0$, and the above expression for ρ into Equation A.29, and using the same small angle approximation as in section A.9, results in

$$\begin{aligned} u &= \frac{v_y \phi'}{R_y - \sqrt{r_c^2 - \phi'^2 R_y^2}}, \\ v &= \frac{v_y \theta'}{R_y - \sqrt{r_c^2 - \phi'^2 R_y^2}}. \end{aligned} \tag{A.35}$$

The deformation is then given by

$$\text{def} = \sqrt{(u_{\phi'} - v_{\theta'})^2 + (v_{\phi'} + u_{\theta'})^2}. \tag{A.36}$$

To assess the effect of the cylinder's curvature on the deformation, I also calculated the deformation induced by a local tangent plane. The distance function associated with the tangent plane whose normal makes an angle τ with the viewing direction and which is located at $(s_\tau r_c, R_y - c_\tau r_c, 0)$ is

$$\begin{aligned} \rho &= \frac{R_y c_\tau - r_c}{s_\theta (s_\phi c_\tau - c_\phi s_\tau)} \\ &\approx \frac{R_y c_\tau - r_c}{c_\tau - \phi' s_\tau} \end{aligned} \tag{A.37}$$

To compute the deformation follow the same steps as above. Compare the lines and dotted lines in Figure 5.1.

A.11 Deformation of the optic flow field of a cylinder rotating about its vertical length axis

Substituting $\Omega = (0, 0, \omega_z)$, and expression A.34 for ρ into Equations A.27 and A.28, and using the same small angle approximation as in section A.9, gives

$$\begin{aligned} u &= -\frac{\omega_z \sqrt{r_c'^2 - \phi'^2}}{1 - \sqrt{r_c'^2 - \phi'^2}}, \\ v &= \frac{\omega_z \theta' \phi'}{1 - \sqrt{r_c'^2 - \phi'^2}}, \end{aligned} \tag{A.38}$$

where $r_c' = r_c / R_y$.

The deformation is again given by Equation A.36.

A.12 Deformation of the optic flow field of a plane translating in a frontoparallel direction

Let a plane translate by $\mathbf{T} = (v_x, 0, 0)$. To compute the resulting velocity field, substitute $\mathbf{T}' = (-v_x, 0, 0), \varpi = 0$, and the expression for ρ given in Equation A.31 (and letting $\omega = 0$ in that equation) into Equations A.29:

$$\begin{aligned} u &= -\frac{v_x}{R_y}(1 - \theta' \tan\sigma) \\ v &= \frac{v_x \theta' \phi'}{R_y}(1 - \theta' \tan\sigma). \end{aligned} \tag{A.39}$$

The partial derivatives with respect to ϕ' and θ' are

$$\begin{aligned} u_{\phi'} &= 0 \\ u_{\theta'} &= \tan\sigma v_x / R_y \\ v_{\phi'} &= \frac{v_x \theta'}{R_y}(1 - \theta' \tan\sigma) \\ v_{\theta'} &= \frac{v_x \phi'}{R_y}(1 - 2\theta' \tan\sigma), \end{aligned} \tag{A.40}$$

so that

$$\begin{aligned} \text{def} &= \left((u_{\phi'} - v_{\theta'})^2 + (v_{\phi'} + u_{\theta'})^2\right)^{1/2} \\ &= \left|\frac{v_x}{R_y}\right| \left(\phi'^2(1 - 2\theta' \tan\sigma)^2 + (\theta'(1 - \theta' \tan\sigma) + \tan\sigma)^2\right)^{1/2} \\ &\approx \left|\frac{v_x}{R_y} \tan\sigma\right|. \end{aligned} \tag{A.41}$$

To compute the deformation for the translating planes in Braunstein & Tittle (1988), I used their average image velocity for v_x/R_y. Thus, the deformation of the plane with average image velocity $6°/s$ simulating a slant of 47° (the first entry in Table 5.2) is def = $6(/s) \times (180/\pi) \times \tan(47°) = .112/s$.

The angle, δ, of the axis of contraction with the horizontal is

$$\begin{aligned} \delta &= \frac{1}{2}\text{atan}\left(\frac{v_{\phi'} + u_{\theta'}}{u_{\phi'} - v_{\theta'}}\right) \\ &= \frac{1}{2}\text{atan}\left(\frac{\theta'(1 - \theta' \tan\sigma) + \tan\sigma}{-\phi'(1 - 2\theta' \tan\sigma)}\right) \\ &\approx \frac{1}{2}\text{atan}\left(\frac{\tan\sigma + \theta'}{-\phi'}\right) \\ &\approx 45°. \end{aligned} \tag{A.42}$$

Thus, the axis of contraction bisects the direction of slant (vertical and downwards) and the direction of $\mathbf{A}_t$ (along the horizontal, in the direction of the negative x-axis).

A.13 Deformation of the optic flow field of a plane rotating about a frontoparallel axis

The velocity field is given in Equation A.33. The partial derivatives of u and v with respect to ϕ' and θ' are

$$\begin{aligned} u_{\phi'} &= 0 \\ u_{\theta'} &= \tan\sigma\omega_z \\ v_{\phi'} &= \omega_z\theta'(1-\theta'\tan\sigma) \\ v_{\theta'} &= \omega_z\phi'(1-2\theta'\tan\sigma), \end{aligned} \tag{A.43}$$

where I used $\omega = 0$. that is, limiting our attention to the case where tilt of the plane is 0° with respect to the vertical. Then,

$$\begin{aligned} \text{def} &= \left((u_{\phi'}-v_{\theta'})^2+(v_{\phi'}+u_{\theta'})^2\right)^{1/2} \\ &= |\omega_z|\left(\phi'^2(1-2\theta'\tan\sigma)^2+(\theta'(1-\theta'\tan\sigma)+\tan\sigma)^2\right)^{1/2} \\ &\approx |\omega_z\tan\sigma|. \end{aligned} \tag{A.44}$$

Thus, the deformation of the plane rotating at $17.2°/s$ having a slant of 22° (the fifth entry in Table 5.2) is def $= 17.2(/s) \times (180/\pi) \times \tan(22°) = .121/s$.

The angle between the axis of contraction and the horizontal is again given by Equation A.42, and is approximately 45°, bisecting the direction of slant and the direction of $\mathbf{A}_t$.

A.14 Deformation of the optic flow field of a plane translating in depth

Let a plane translate by $\mathbf{T} = (0, v_y, 0)$. To compute the resulting velocity field, substitute $\mathbf{T}' = (0, -v_y, 0), \varpi = 0$, and the expression for ρ given in Equation A.31 (and letting $\omega = 0$ in that equation) into Equations A.29:

$$\begin{aligned} u &= v'_y\phi'(1-\theta'\tan\sigma) \\ v &= v'_y\theta'(1-\theta'\tan\sigma), \end{aligned} \tag{A.45}$$

where $v'_y = v_y/R_y$. The partial derivatives with respect to ϕ' and θ' are

$$\begin{aligned} u_{\phi'} &= v'_y(1-\theta'\tan\sigma) \\ u_{\theta'} &= -v'_y\phi'\tan\sigma \\ v_{\phi'} &= 0 \\ v_{\theta'} &= v'_y(1-2\theta'\tan\sigma), \end{aligned} \tag{A.46}$$

so that

$$\begin{aligned} \text{def} &= \left((u_{\phi'}-v_{\theta'})^2+(v_{\phi'}+u_{\theta'})^2\right)^{1/2} \\ &= |v'_y|\left((1-\theta'\tan\sigma-(1-2\theta'\tan\sigma))^2+(\phi'\tan\sigma)^2\right)^{1/2} \\ &\approx |v'_y\tan\sigma|(\theta'^2+\phi'^2)^{1/2}. \end{aligned} \tag{A.47}$$

The angle, δ, of the axis of contraction with the horizontal is

$$\begin{aligned}\delta &= \frac{1}{2}\text{atan}\left(\frac{v_{\phi'} + u_{\theta'}}{u_{\phi'} - v_{\theta'}}\right) \\ &= \frac{1}{2}\text{atan}\left(\frac{-\phi' \tan\sigma}{1 - \theta' \tan\sigma - (1 - 2\theta' \tan\sigma)}\right) \\ &\approx \frac{1}{2}\text{atan}\left(\frac{-\phi'}{\theta'}\right).\end{aligned} \tag{A.48}$$

Note that the amount of deformation for translation in depth is much smaller than for translation in a frontoparallel direction (compare Equations A.48 and A.41).

Appendix B

Discriminating ellipsoids and ellipses: Raw data

Tables B.1 through B.12 describe the families of ellipsoids $a_{11}x^2 + a_{22}y^2 + a_{33}z^2 = 1$ used in Experiment 6.1 in which the subject had to indicate whether the perceived object was "flat" or "not flat." The tables also list the discrimination measures and some 2-D measures, as well as the results of the experiments. In Tables B.1 through B.8, the angular velocity ($\dot{\theta}$) used to generate the stimuli was used to compute the discrimination measures; Tables B.11 and B.12 show the discrimination measures if the estimated θ is used (B.11 corresponds to B.1, and B.12 to B.3; see section 6.2 for estimating angular velocity).

The following are detailed explanations of the column headings of Tables B.1 through B.12.

- a_{11}: x^2 coefficient.
- a_{22}: y^2 coefficient.
- R_V: Range of rotation about the (vertical) axis of rotation.
- θ: amount of rotation in degrees between successive views.
- **Est** θ: θ estimated from the sequence of contours assuming correspondence between successive views (section 6.2).
- **% e:** percentage error in the estimate of θ.
- **Area:** each contour circumscribes an elliptic region in the image; **Area** is the sum of the area of these regions summed over all views.
- **Circf:** each contour circumscribes an elliptic region in the image; **Circf** is the sum of the circumference of these regions summed over all views.
- **Contr:** consider a horizontal cross section of the projection; as the ellipsoid rotates about the vertical, the length of this section changes, attaining some maximal and minimal value. The amount of contraction of a section is defined as (max-min)/(max+min). **Contr** is the mean contraction of all horizontal cross sections. The values given are obtained with 5% noise. The standard deviation is about 10% for the ellipse and flat ellipsoids having a contraction of more than .05; standard deviation increases with decreasing contraction, reaching 100% for contractions less than .03.
- **Motion:** Radius of curvature in the direction of motion summed along the contour and over all views. Value in square brackets is the increase of this measure with respect to the flat ellipse.
- **Radial:** Radius of curvature of the normal section summed along the contour and over all views.
- **Nonrigid:** Nonrigidity discrimination measure. Value in square brackets is the increase of this measure with respect to the flat ellipse.

- **Rim:** Rim prediction measure. Value in square brackets is the increase of this measure with respect to the flat ellipse.
- **Contour:** Contour prediction measure.
- **Tangent:** Tangent prediction measure.
- **Gauss:** Radius of Gaussian curvature summed along the contour and over all views.
- **Mean:** Radius of the mean curvature summed along the contour and over all views.
- **Max V:** Maximal velocity in the image in the horizontal direction during the entire sequence.
- **Mean V:** Mean velocity in the image in the horizontal direction; standard deviation is about 60% of the mean velocity.
- **Max Ext:** Maximal horizonal extent of the occluding contours during the entire sequence.
- **Flat:** Number of "flat" responses.
- **¬Flat:** Number of "not flat" responses.
- d': d' and significance interval were computed using formulae given in Marascuilo (1970).

Figure B.1 shows the psychometric curves associated with the motion, nonrigidity, rim prediction, and tangent plane prediction measures for experiments B-1 through B-8, each computed using the actual and estimated angular velocity:

(a) motion discrimination measure computed using the actual angular velocity.
(b) motion discrimination measure computed using the estimated angular velocity.
(c) nonrigidity discrimination measure computed using the actual angular velocity.
(d) nonrigidity discrimination measure computed using the estimated angular velocity.
(e) rim prediction discrimination measure computed using the actual angular velocity.
(f) rim prediction discrimination measure computed using the estimated angular velocity.
(g) tangent plane prediction discrimination measure computed using the actual angular velocity.
(h) tangent plane prediction discrimination measure computed using the estimated angular velocity.
(i) contraction discrimination measure.

Table B.1 Contour sequence. Used actual θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-20.0,40.0]	4.0	4.1	5	459	309	0.111
1	0.750	0.048	[-20.4,40.8]	4.1	3.6	10	325	272	0.082
2	1.000	0.068	[-20.6,41.2]	4.1	3.7	8	374	284	0.083
3	2.000	0.097	[-20.9,41.7]	4.2	3.8	6	522	326	0.095
4	1.250	0.139	[-21.2,42.5]	4.2	3.5	9	414	295	0.067
5	2.000	0.198	[-21.8,43.5]	4.4	3.6	9	513	323	0.072
6	1.500	0.283	[-22.6,45.1]	4.5	3.1	18	444	302	0.039
7	1.500	0.405	[-23.7,47.4]	4.7	2.7	30	440	300	0.011
8	2.000	0.578	[-25.5,50.9]	5.1	2.8	29	490	313	0.018
9	2.000	0.826	[-28.1,56.3]	5.6	3.5	55	479	309	0.026

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.5 [1.0]	0.9	1.1 [1.0]	0.8 [1.0]	0.8	1.0 [1.0]	0.7	1.7
1	0.5 [1.2]	0.7	1.4 [1.2]	0.7 [1.0]	0.5	0.7 [0.8]	0.3	0.9
2	0.6 [1.4]	0.9	1.5 [1.3]	0.9 [1.2]	0.6	0.8 [0.9]	0.3	1.1
3	0.8 [1.8]	1.4	1.5 [1.3]	1.1 [1.4]	0.8	1.3 [1.3]	2.1	5.8
4	1.0 [2.2]	1.3	2.0 [1.8]	1.4 [1.9]	0.7	1.2 [1.2]	0.7	2.2
5	1.2 [2.6]	1.9	2.1 [1.8]	1.8 [2.3]	0.9	1.4 [1.5]	2.5	7.4
6	1.6 [3.5]	2.2	2.9 [2.6]	2.6 [3.4]	1.0	1.8 [1.9]	1.4	4.3
7	2.1 [4.6]	2.8	3.6 [3.2]	3.7 [4.9]	1.3	2.2 [2.2]	1.8	5.3
8	2.5 [5.6]	3.6	3.8 [3.3]	3.4 [4.4]	1.1	2.2 [2.2]	4.5	12.6
9	3.2 [7.1]	4.5	4.5 [3.9]	5.2 [6.9]	1.6	2.9 [3.0]	5.6	15.6

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]
0	100	100	100	68	22	
1	98	96	85	6	4	0.44 ± 0.84
2	98	96	90	6	4	0.44 ± 0.84
3	97	97	107	6	4	0.44 ± 0.84
4	96	94	94	3	7	1.22 ± 0.86*
5	95	93	105	4	6	0.94 ± 0.84*
6	94	90	97	1	9	1.97 ± 1.10*
7	91	87	96	0	10	5.89 ± 0.28*
8	89	83	101	0	10	5.89 ± 0.28*
9	85	79	100	0	10	5.89 ± 0.28*

[1] 52% of the total number of responses were "flat", 48% were "¬flat".
[2] * indicates significance at 5%.

Table B.2 Silhouette sequence. Used actual θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-20.0,40.0]	4.0	4.1	5	459	309	0.111
1	0.750	0.048	[-20.4,40.8]	4.1	3.6	10	325	272	0.082
2	1.000	0.068	[-20.6,41.2]	4.1	3.7	8	374	284	0.083
3	2.000	0.097	[-20.9,41.7]	4.2	3.8	6	522	326	0.095
4	1.250	0.139	[-21.2,42.5]	4.2	3.5	9	414	295	0.067
5	2.000	0.198	[-21.8,43.5]	4.4	3.6	9	513	323	0.072
6	1.500	0.283	[-22.6,45.1]	4.5	3.1	18	444	302	0.039
7	1.500	0.405	[-23.7,47.4]	4.7	2.7	30	440	300	0.011
8	2.000	0.578	[-25.5,50.9]	5.1	2.8	29	490	313	0.018

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.5 [1.0]	0.9	1.1 [1.0]	0.8 [1.0]	0.8	1.0 [1.0]	0.7	1.7
1	0.5 [1.2]	0.7	1.4 [1.2]	0.7 [1.0]	0.5	0.7 [0.8]	0.3	0.9
2	0.6 [1.4]	0.9	1.5 [1.3]	0.9 [1.2]	0.6	0.8 [0.9]	0.3	1.1
3	0.8 [1.8]	1.4	1.5 [1.3]	1.1 [1.4]	0.8	1.3 [1.3]	2.1	5.8
4	1.0 [2.2]	1.3	2.0 [1.8]	1.4 [1.9]	0.7	1.2 [1.2]	0.7	2.2
5	1.2 [2.6]	1.9	2.1 [1.8]	1.8 [2.3]	0.9	1.4 [1.5]	2.5	7.4
6	1.6 [3.5]	2.2	2.9 [2.6]	2.6 [3.4]	1.0	1.8 [1.9]	1.4	4.3
7	2.1 [4.6]	2.8	3.6 [3.2]	3.7 [4.9]	1.3	2.2 [2.2]	1.8	5.3
8	2.5 [5.6]	3.6	3.8 [3.3]	3.4 [4.4]	1.1	2.2 [2.2]	4.5	12.6

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]	
0	100	100	100	55	25		
1	98	96	85	10	0	-4.71	± 0.29*
2	98	96	90	6	4	0.24	± 0.84
3	97	97	107	7	3	-0.04	± 0.87
4	96	94	94	1	9	1.77	± 1.10*
5	95	93	105	3	7	1.01	± 0.87*
6	94	90	97	0	10	5.69	± 0.29*
7	91	87	96	0	10	5.69	± 0.29*
8	89	83	101	0	10	5.69	± 0.29*

[1] 51% of the total number of responses were "flat", 49% were "¬flat".
[2] * indicates significance at 5%.

Table B.3 Contour sequence. Used actual θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.25 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-20.0,40.0]	4.0	4.0	4	562	340	0.111
1	0.750	0.048	[-20.6,41.3]	4.1	3.6	8	396	290	0.084
2	1.000	0.068	[-20.9,41.8]	4.2	3.7	7	453	306	0.086
3	2.000	0.097	[-21.3,42.6]	4.3	3.8	5	630	360	0.099
4	1.250	0.139	[-21.9,43.7]	4.4	3.5	9	497	318	0.072
5	2.000	0.198	[-22.7,45.4]	4.5	3.7	7	612	353	0.081
6	1.500	0.283	[-23.9,47.8]	4.8	3.2	14	525	325	0.049
7	1.500	0.405	[-25.8,51.5]	5.2	3.0	26	513	320	0.024
8	2.000	0.578	[-28.6,57.2]	5.7	3.2	23	563	334	0.034

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.5 [1.0]	1.2	1.2 [1.0]	0.9 [1.0]	1.0	1.3 [1.0]	1.2	2.6
1	0.6 [1.2]	0.9	1.5 [1.2]	0.9 [1.0]	0.6	0.9 [0.7]	0.3	1.0
2	0.8 [1.5]	1.1	1.6 [1.3]	1.1 [1.2]	0.7	1.1 [0.9]	0.4	1.4
3	1.0 [1.9]	2.1	1.6 [1.3]	1.4 [1.6]	1.0	1.7 [1.4]	4.8	9.8
4	1.2 [2.4]	1.8	2.2 [1.8]	1.9 [2.1]	0.9	1.5 [1.2]	1.0	2.9
5	1.5 [2.8]	2.6	2.2 [1.8]	2.4 [2.7]	1.2	2.0 [1.6]	4.7	11.5
6	2.0 [3.8]	3.0	3.0 [2.5]	3.6 [4.0]	1.3	2.5 [2.0]	2.2	6.0
7	2.6 [4.9]	3.8	3.7 [3.0]	3.5 [3.9]	1.2	2.4 [1.9]	2.7	7.4
8	3.2 [6.0]	5.0	3.8 [3.1]	5.3 [5.9]	1.7	3.2 [2.5]	7.5	19.0

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]	
0	100	100	100	58	22		
1	98	94	82	7	3	0.07	± 0.87
2	97	94	87	4	6	0.85	± 0.84*
3	96	96	106	8	2	-0.24	± 0.93
4	95	91	92	3	7	1.12	± 0.87*
5	93	91	104	7	3	0.07	± 0.87
6	91	86	94	0	10	5.80	± 0.29*
7	88	81	92	0	10	5.80	± 0.29*
8	85	77	96	1	9	1.88	± 1.10*

[1] 55% of the total number of responses were "flat", 45% were "¬flat".
[2] * indicates significance at 5%.

Table B.4 Contour sequence. Used actual θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-35.3,24.7]	4.0	4.1	5	460	309	0.106
1	0.750	0.048	[-34.0,27.2]	4.1	3.8	10	326	270	0.071
2	1.000	0.068	[-32.5,29.3]	4.1	3.8	8	376	283	0.065
3	2.000	0.097	[-31.3,31.3]	4.2	3.8	7	526	325	0.065
4	1.250	0.139	[-30.3,33.4]	4.2	3.5	12	416	293	0.043
5	2.000	0.198	[-29.7,35.6]	4.4	3.7	8	517	322	0.045
6	1.500	0.283	[-29.4,38.2]	4.5	3.1	16	446	300	0.027
7	1.500	0.405	[-29.6,41.5]	4.7	2.7	31	440	298	0.009

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.5 [1.0]	0.9	1.1 [1.0]	0.8 [1.0]	0.8	1.0 [1.0]	0.7	1.8
1	0.6 [1.2]	0.8	1.4 [1.2]	0.7 [1.0]	0.5	0.7 [0.7]	0.3	0.8
2	0.6 [1.3]	0.8	1.5 [1.3]	0.9 [1.1]	0.6	0.8 [0.8]	0.3	1.1
3	0.8 [1.7]	1.5	1.5 [1.3]	1.1 [1.4]	0.8	1.3 [1.3]	2.5	6.2
4	1.0 [2.0]	1.4	2.0 [1.8]	1.4 [1.8]	0.7	1.1 [1.2]	0.7	2.1
5	1.2 [2.4]	1.9	2.0 [1.8]	1.7 [2.3]	0.9	1.4 [1.4]	2.7	7.6
6	1.6 [3.2]	2.2	2.9 [2.5]	2.5 [3.3]	1.0	1.8 [1.8]	1.4	4.1
7	2.1 [4.2]	2.8	3.6 [3.2]	3.6 [4.8]	1.3	2.1 [2.1]	1.8	5.2

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'^2	
0	100	100	100	42	28		
1	98	96	85	6	4	0.00	± 0.84
2	98	96	89	3	7	0.78	± 0.87
3	97	97	106	8	2	-0.59	± 0.93
4	96	94	94	5	5	0.25	± 0.83
5	95	94	105	6	4	0.00	± 0.84
6	94	91	97	0	10	5.45	± 0.30*
7	91	88	96	2	8	1.09	± 0.93*

[1] 51% of the total number of responses were "flat", 49% were "¬flat".
[2] * indicates significance at 5%.

Table B.5 Contour sequence. Used actual θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-35.3,24.7]	4.0	4.1	5	460	309	0.106
1	1.000	0.048	[-34.0,27.2]	4.1	3.8	8	376	283	0.077
2	1.250	0.068	[-32.5,29.3]	4.1	3.9	7	419	295	0.070
3	2.000	0.097	[-31.3,31.3]	4.2	3.8	7	526	325	0.065
4	1.500	0.139	[-30.3,33.4]	4.2	3.7	9	454	304	0.048
5	1.500	0.198	[-29.7,35.6]	4.4	3.5	11	451	302	0.035
6	2.000	0.283	[-29.4,38.2]	4.5	3.4	10	510	319	0.042
7	2.000	0.405	[-29.6,41.5]	4.7	3.1	17	501	316	0.031

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.5 [1.0]	0.9	1.1 [1.0]	0.8 [1.0]	0.8	1.0 [1.0]	0.7	1.8
1	0.6 [1.2]	0.9	1.3 [1.2]	0.7 [1.0]	0.6	0.8 [0.8]	0.3	1.0
2	0.6 [1.3]	0.9	1.4 [1.2]	0.9 [1.1]	0.6	0.9 [0.9]	0.4	1.4
3	0.8 [1.7]	1.5	1.5 [1.3]	1.1 [1.4]	0.8	1.3 [1.3]	2.5	6.2
4	1.0 [2.0]	1.4	1.8 [1.6]	1.4 [1.8]	0.8	1.2 [1.2]	0.9	3.0
5	1.2 [2.5]	1.7	2.3 [2.0]	1.8 [2.3]	0.8	1.3 [1.3]	1.1	3.6
6	1.5 [3.1]	2.3	2.5 [2.2]	2.4 [3.1]	1.0	1.9 [2.0]	3.0	8.5
7	2.0 [4.0]	2.9	3.1 [2.7]	3.4 [4.5]	1.3	2.2 [2.2]	3.7	10.4

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]	
0	100	100	100	34	36		
1	98	97	90	7	3	-0.56	± 0.87
2	98	97	94	6	4	-0.29	± 0.84
3	97	97	106	1	9	1.25	± 1.10*
4	96	95	98	4	6	0.22	± 0.84
5	95	93	97	5	5	-0.04	± 0.83
6	94	91	104	3	7	0.49	± 0.87
7	91	88	103	0	10	5.16	± 0.29*

[1] 43% of the total number of responses were "flat", 57% were "¬flat".
[2] * indicates significance at 5%.

Table B.6 Contour sequence. Used actual θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.600	0.000	[-35.3,21.8]	3.8	3.8	5	476	313	0.106
1	1.600	0.068	[-34.3,25.2]	4.0	3.7	8	473	310	0.079
2	1.600	0.139	[-33.5,28.6]	4.1	3.7	9	468	308	0.061
3	1.700	0.198	[-32.7,31.8]	4.3	3.5	11	478	310	0.046
4	1.700	0.283	[-32.2,35.4]	4.5	3.3	15	473	308	0.031
5	1.800	0.405	[-32.3,38.8]	4.7	3.1	22	478	309	0.020
6	2.000	0.578	[-33.2,43.2]	5.1	2.7	30	490	312	0.011

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.5 [1.0]	1.0	1.1 [1.0]	0.7 [1.0]	0.7	1.0 [1.0]	0.9	2.1
1	0.7 [1.4]	1.1	1.4 [1.2]	1.0 [1.5]	0.8	1.2 [1.2]	0.9	2.7
2	0.9 [1.9]	1.4	1.8 [1.6]	1.2 [1.7]	0.7	1.1 [1.1]	1.1	3.2
3	1.2 [2.5]	1.9	2.1 [1.9]	1.7 [2.4]	0.8	1.4 [1.4]	1.7	5.0
4	1.6 [3.2]	2.2	2.7 [2.4]	2.4 [3.4]	1.0	1.6 [1.7]	1.9	5.7
5	2.0 [4.1]	2.8	3.3 [2.9]	3.5 [4.9]	1.3	2.1 [2.2]	2.7	7.9
6	2.5 [5.2]	3.6	3.8 [3.3]	3.3 [4.7]	1.1	2.1 [2.2]	4.4	12.5

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]
0	100	100	100	33	27	
1	99	98	99	4	6	0.38 ± 0.85
2	99	97	98	6	4	-0.13 ± 0.85
3	99	97	99	5	5	0.13 ± 0.84
4	98	95	99	7	3	-0.40 ± 0.88
5	96	92	99	3	7	0.65 ± 0.88
6	93	88	100	1	9	1.41 ± 1.11*

[1] 49% of the total number of responses were "flat", 51% were "¬flat".
[2] * indicates significance at 5%.

Table B.7 Contour sequence. Used actual θ in computing discrimination measures. Number of views: 9; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-12.0,24.0]	4.0	3.9	8	286	182	0.030
1	0.750	0.048	[-12.2,24.5]	4.1	3.6	11	201	157	0.025
2	1.000	0.068	[-12.3,24.7]	4.1	3.8	7	231	165	0.024
3	2.000	0.097	[-12.5,25.0]	4.2	3.7	5	323	193	0.031
4	1.250	0.139	[-12.7,25.4]	4.2	3.4	4	254	172	0.017
5	2.000	0.198	[-13.0,26.1]	4.3	3.4	7	316	191	0.024
6	1.500	0.283	[-13.5,27.0]	4.5	2.9	11	271	177	0.012
7	1.500	0.405	[-14.2,28.3]	4.7	2.0	25	265	175	0.001
8	2.000	0.578	[-15.1,30.3]	5.0	2.0	21	295	184	0.005

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.2 [1.0]	0.4	1.1 [1.0]	0.3 [1.0]	0.3	0.5 [1.0]	0.3	0.9
1	0.3 [1.3]	0.4	1.1 [1.0]	0.3 [1.0]	0.2	0.3 [0.6]	0.1	0.4
2	0.4 [1.8]	0.5	1.2 [1.0]	0.3 [1.1]	0.2	0.3 [0.7]	0.2	0.6
3	0.4 [1.9]	0.7	1.2 [1.1]	0.4 [1.5]	0.3	0.6 [1.3]	1.4	3.7
4	0.5 [2.3]	0.7	1.4 [1.2]	0.5 [1.8]	0.3	0.5 [1.1]	0.3	1.0
5	0.6 [2.8]	1.0	1.4 [1.3]	0.7 [2.3]	0.4	0.7 [1.5]	1.6	4.5
6	0.8 [3.7]	1.2	1.8 [1.6]	1.0 [3.3]	0.4	0.7 [1.5]	0.7	2.3
7	1.1 [4.9]	1.4	2.2 [1.9]	1.4 [4.7]	0.5	0.8 [1.9]	0.9	2.7
8	1.4 [6.2]	2.0	2.3 [2.0]	1.4 [4.7]	0.4	0.9 [2.0]	2.5	7.2

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]
0	100	100	100	49	31	
1	98	98	77	9	1	-1.00 ± 1.10
2	98	97	84	8	2	-0.56 ± 0.93
3	97	97	110	9	1	-1.00 ± 1.10
4	96	95	91	5	5	0.29 ± 0.83
5	95	94	108	3	7	0.81 ± 0.86
6	93	92	95	2	8	1.13 ± 0.93*
7	91	89	93	2	8	1.13 ± 0.93*
8	88	86	101	3	7	0.81 ± 0.86

[1] 56% of the total number of responses were "flat", 44% were "¬flat".
[2] * indicates significance at 5%.

Table B.8 Contour sequence. Used actual θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.17 throughout. All ellipsoids were rotated 45° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-20.0,40.0]	4.0	4.0	4	325	259	0.111
1	0.750	0.048	[-20.3,40.5]	4.1	3.9	8	230	235	0.097
2	1.000	0.068	[-20.4,40.7]	4.1	3.9	6	265	244	0.098
3	2.000	0.097	[-20.5,41.0]	4.1	4.0	5	373	272	0.104
4	1.250	0.139	[-20.7,41.5]	4.1	3.8	7	296	251	0.089
5	2.000	0.198	[-21.1,42.1]	4.2	3.8	6	371	272	0.090
6	1.500	0.283	[-21.6,43.1]	4.3	3.5	8	322	258	0.073
7	1.500	0.405	[-22.3,44.5]	4.5	3.4	10	322	258	0.058
8	2.000	0.578	[-23.3,46.7]	4.7	3.4	11	366	270	0.062

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.2 [1.0]	0.3	0.8 [1.0]	0.4 [1.0]	0.5	0.5 [1.0]	0.1	0.7
1	0.2 [1.1]	0.3	0.7 [1.0]	0.3 [0.8]	0.3	0.3 [0.7]	0.1	0.4
2	0.3 [1.2]	0.3	0.8 [1.0]	0.4 [0.9]	0.4	0.4 [0.8]	0.1	0.5
3	0.3 [1.6]	0.5	0.8 [1.1]	0.5 [1.1]	0.5	0.6 [1.2]	0.5	2.0
4	0.4 [1.8]	0.5	0.9 [1.2]	0.5 [1.3]	0.4	0.5 [1.1]	0.2	0.8
5	0.5 [2.1]	0.6	1.0 [1.3]	0.7 [1.6]	0.6	0.6 [1.3]	0.5	2.4
6	0.6 [3.0]	0.8	1.2 [1.6]	0.9 [2.2]	0.5	0.7 [1.5]	0.4	1.6
7	0.8 [3.9]	1.0	1.5 [1.9]	1.3 [3.0]	0.6	0.8 [1.8]	0.5	2.0
8	1.0 [4.9]	1.3	1.6 [2.2]	1.7 [4.2]	0.8	1.0 [2.2]	1.0	4.3

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]	
0	100	100	100	56	24		
1	99	96	82	10	0	-4.68	± 0.29*
2	99	97	88	8	2	-0.32	± 0.93
3	99	99	109	4	6	0.78	± 0.84
4	99	96	94	6	4	0.27	± 0.84
5	98	98	109	4	6	0.78	± 0.84
6	98	94	99	2	8	1.37	± 0.93*
7	97	92	100	3	7	1.05	± 0.87*
8	96	92	108	1	9	1.81	± 1.10*

[1] 59% of the total number of responses were "flat", 41% were "¬flat".
[2] * indicates significance at 5%.

Table B.9 Contour sequence. Number of views: 15; θ is the angle in degrees between views. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about an axis slanted by −22.5°.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	**Flat**[1]	**¬Flat**		d'^2
0	1.600	0.000	[-35.3,21.8]	3.81	61	9		
1	1.600	0.068	[-34.3,25.2]	3.97	8	2	0.29	± 0.96
2	1.600	0.139	[-33.5,28.6]	4.14	9	1	-0.15	± 1.12
3	1.700	0.198	[-32.7,31.8]	4.30	7	3	0.61	± 0.90
4	1.700	0.283	[-32.2,35.4]	4.51	6	4	0.88	± 0.87*
5	1.800	0.405	[-32.3,38.8]	4.74	2	8	1.97	± 0.96*
6	2.000	0.578	[-33.2,43.2]	5.09	3	7	1.66	± 0.90*
7	2.000	0.826	[-35.2,49.2]	5.63	0	10	6.33	± 0.37*

[1] 69% of the total number of responses were "flat", 31% were "¬flat".
[2] * indicates significance at 5%.

Table B.10 Contour sequence. Number of views: 15; θ is the angle in degrees between views. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about an axis slanted by 22.5°.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	**Flat**[1]	**¬Flat**		d'^2
0	1.600	0.000	[-35.3,21.8]	3.81	53	17		
1	1.600	0.068	[-34.3,25.2]	3.97	9	1	-0.58	± 1.11
2	1.600	0.139	[-33.5,28.6]	4.14	6	4	0.44	± 0.85
3	1.700	0.198	[-32.7,31.8]	4.30	4	6	0.95	± 0.85*
4	1.700	0.283	[-32.2,35.4]	4.51	4	6	0.95	± 0.85*
5	1.800	0.405	[-32.3,38.8]	4.74	3	7	1.22	± 0.88*
6	2.000	0.578	[-33.2,43.2]	5.09	0	10	5.90	± 0.32*
7	2.000	0.826	[-35.2,49.2]	5.63	0	10	5.90	± 0.32*

[1] 56% of the total number of responses were "flat", 44% were "¬flat".
[2] * indicates significance at 5%.

Table B.11 Contour sequence. Used estimated θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.17 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-20.0,40.0]	4.0	4.1	5	459	309	0.111
1	0.750	0.048	[-20.4,40.8]	4.1	3.6	10	325	272	0.082
2	1.000	0.068	[-20.6,41.2]	4.1	3.7	8	374	284	0.083
3	2.000	0.097	[-20.9,41.7]	4.2	3.8	6	522	326	0.095
4	1.250	0.139	[-21.2,42.5]	4.2	3.5	9	414	295	0.067
5	2.000	0.198	[-21.8,43.5]	4.4	3.6	9	513	323	0.072
6	1.500	0.283	[-22.6,45.1]	4.5	3.1	18	444	302	0.039
7	1.500	0.405	[-23.7,47.4]	4.7	2.7	30	440	300	0.011
8	2.000	0.578	[-25.5,50.9]	5.1	2.8	29	490	313	0.018
9	2.000	0.826	[-28.1,56.3]	5.6	3.5	55	479	309	0.026

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.5 [1.0]	0.9	1.1 [1.0]	0.8 [1.0]	0.8	1.0 [1.0]	0.7	1.8
1	0.4 [0.9]	0.6	1.2 [1.1]	0.6 [0.8]	0.5	0.8 [0.8]	0.2	0.6
2	0.5 [1.0]	0.7	1.3 [1.1]	0.7 [0.9]	0.6	0.9 [0.9]	0.3	0.8
3	0.6 [1.3]	1.1	1.3 [1.1]	0.8 [1.1]	0.8	1.3 [1.4]	2.0	4.8
4	0.7 [1.5]	1.0	1.5 [1.3]	1.0 [1.3]	0.7	1.3 [1.3]	0.5	1.5
5	0.8 [1.6]	1.4	1.5 [1.3]	1.1 [1.4]	0.9	1.5 [1.5]	2.1	5.2
6	1.3 [2.7]	2.0	2.2 [2.0]	1.6 [2.1]	0.8	2.4 [2.5]	1.4	3.0
7	2.1 [4.5]	3.2	3.4 [3.0]	2.7 [3.5]	0.9	3.5 [3.6]	2.2	4.3
8	2.5 [5.3]	4.1	3.5 [3.1]	2.4 [3.1]	0.9	3.7 [3.8]	5.7	9.8
9	3.9 [8.2]	5.8	5.0 [4.4]	4.4 [5.7]	1.1	4.6 [4.7]	7.2	13.7

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]
0	100	100	100	68	22	
1	98	96	85	6	4	0.44 ± 0.84
2	98	96	90	6	4	0.44 ± 0.84
3	97	97	107	6	4	0.44 ± 0.84
4	96	94	94	3	7	1.22 ± 0.86*
5	95	93	105	4	6	0.94 ± 0.84*
6	94	90	97	1	9	1.97 ± 1.10*
7	91	87	96	0	10	5.89 ± 0.28*
8	89	83	101	0	10	5.89 ± 0.28*
9	85	79	100	0	10	5.89 ± 0.28*

[1] 52% of the total number of responses were "flat", 48% were "¬flat".
[2] * indicates significance at 5%.

Table B.12 Contour sequence. Used estimated θ in computing discrimination measures. Number of views: 15; θ is the angle in degrees between views. Noise is 5%. a_{33} equals 0.25 throughout. All ellipsoids were rotated 30° about the horizontal before being rotated about a vertical axis. Predicted and actual rims, contours, and tangent planes are compared during 20° of rotation.

level	$1/a_{11}$	$1/a_{22}$	R_V	θ	Est θ	%e	Area	Circf	Contr
0	1.500	0.000	[-20.0,40.0]	4.0	4.0	4	562	340	0.111
1	0.750	0.048	[-20.6,41.3]	4.1	3.6	8	396	290	0.084
2	1.000	0.068	[-20.9,41.8]	4.2	3.7	7	453	306	0.086
3	2.000	0.097	[-21.3,42.6]	4.3	3.8	5	630	360	0.099
4	1.250	0.139	[-21.9,43.7]	4.4	3.5	9	497	318	0.072
5	2.000	0.198	[-22.7,45.4]	4.5	3.7	7	612	353	0.081
6	1.500	0.283	[-23.9,47.8]	4.8	3.2	14	525	325	0.049
7	1.500	0.405	[-25.8,51.5]	5.2	3.0	26	513	320	0.024
8	2.000	0.578	[-28.6,57.2]	5.7	3.2	23	563	334	0.034

level	Motion	Radial	Nonrigid	Rim	Contour	Tangent	Gauss	Mean
0	0.5 [1.0]	1.2	1.2 [1.0]	0.9 [1.0]	1.0	1.3 [1.0]	1.2	2.6
1	0.5 [0.9]	0.7	1.2 [1.0]	0.7 [0.8]	0.6	1.0 [0.8]	0.2	0.7
2	0.6 [1.1]	0.9	1.3 [1.1]	0.8 [0.9]	0.7	1.1 [0.9]	0.3	0.9
3	0.7 [1.4]	1.7	1.3 [1.1]	1.0 [1.1]	1.0	1.8 [1.4]	4.5	7.8
4	0.8 [1.6]	1.4	1.6 [1.3]	1.2 [1.3]	0.8	1.6 [1.3]	0.8	1.9
5	0.9 [1.7]	1.8	1.6 [1.3]	1.4 [1.5]	1.1	1.9 [1.5]	3.7	7.8
6	1.4 [2.6]	2.5	2.2 [1.8]	2.0 [2.2]	1.0	3.2 [2.5]	2.0	4.0
7	2.2 [4.3]	3.7	3.1 [2.6]	2.3 [2.5]	0.9	3.7 [2.9]	2.9	5.3
8	2.6 [5.0]	4.9	3.1 [2.6]	3.1 [3.5]	1.3	5.1 [4.0]	8.3	12.8

level	Max V	Mean V	Max Ext	Flat[1]	¬Flat	d'[2]	
0	100	100	100	58	22		
1	98	94	82	7	3	0.07	± 0.87
2	97	94	87	4	6	0.85	± 0.84*
3	96	96	106	8	2	-0.24	± 0.93
4	95	91	92	3	7	1.12	± 0.87*
5	93	91	104	7	3	0.07	± 0.87
6	91	86	94	0	10	5.80	± 0.29*
7	88	81	92	0	10	5.80	± 0.29*
8	85	77	96	1	9	1.88	± 1.10*

[1] 55% of the total number of responses were "flat", 45% were "¬flat".
[2] * indicates significance at 5%.

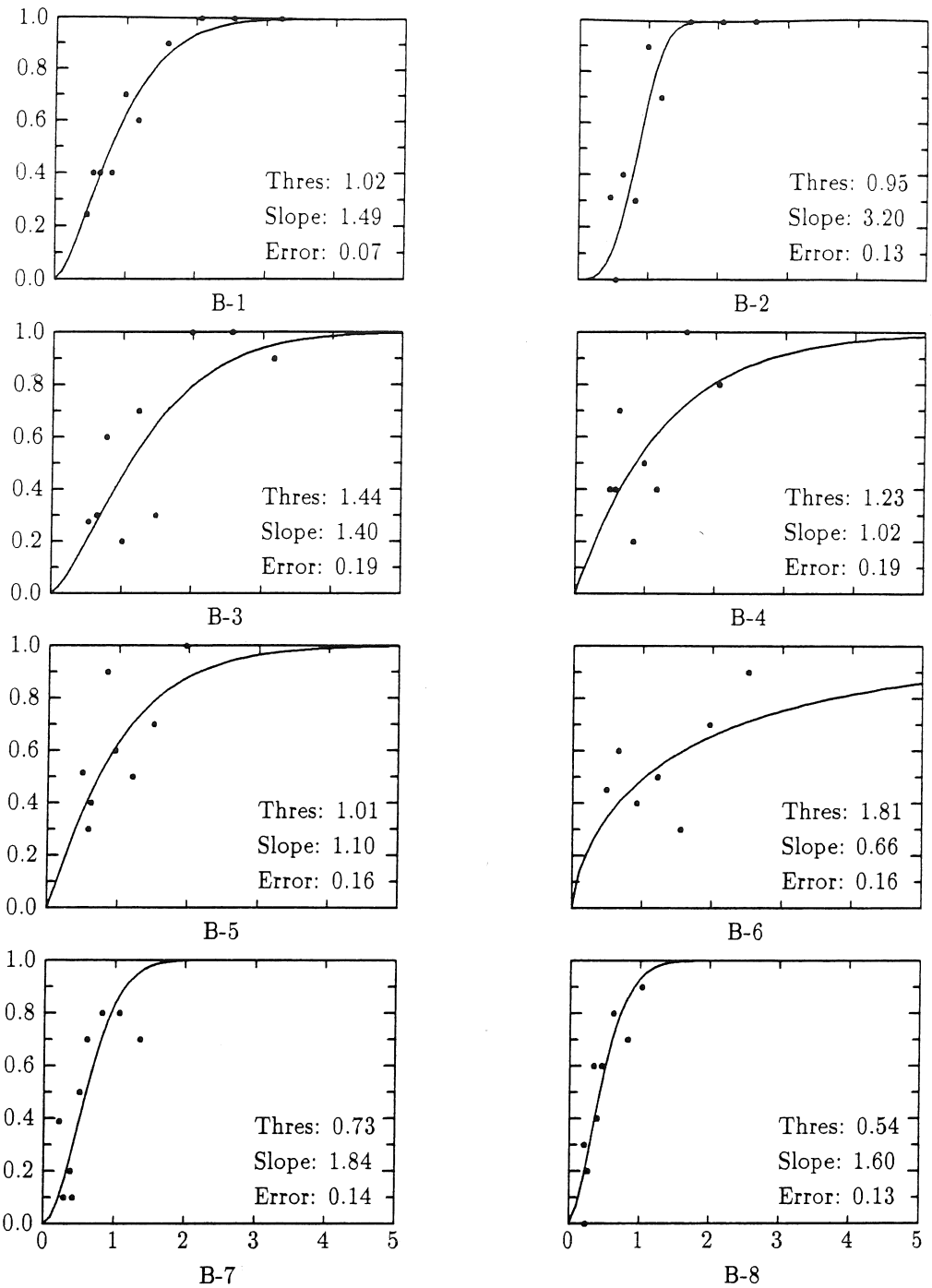


Figure B.1a. Psychometric functions with the *motion* discrimination measure as the independent variable. Used the *actual* angular velocity. Measures were computed using 5% noise. Thresholds are absolute thresholds. (See column Motion in Tables B-1 through B-8.)

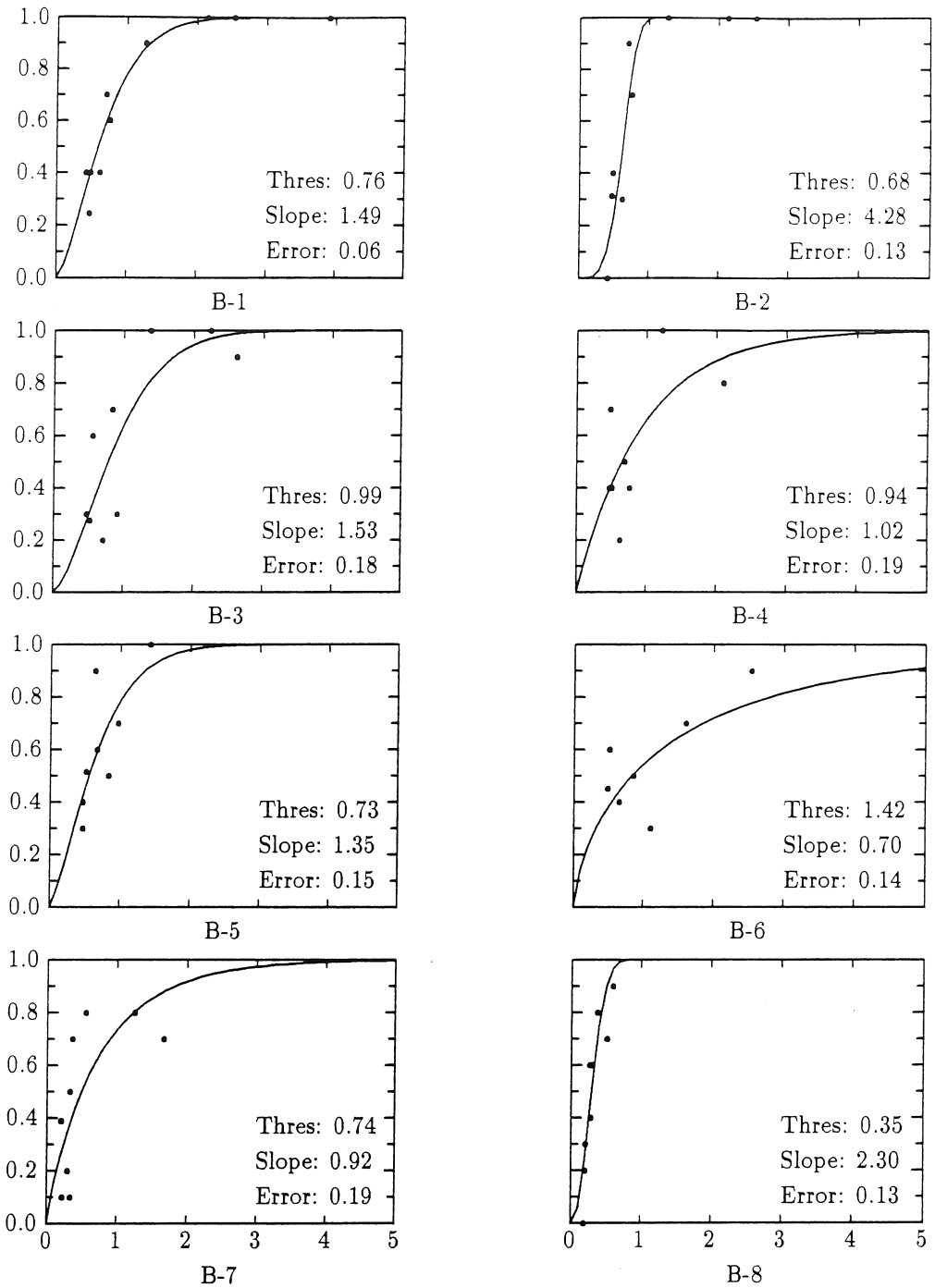


Figure B.1b. Psychometric functions with the *motion* discrimination measure as the independent variable. Measures were computed using 5% noise and the *estimated* angular velocity. Thresholds are absolute thresholds. (See column Motion in Tables B-1 through B-8.)

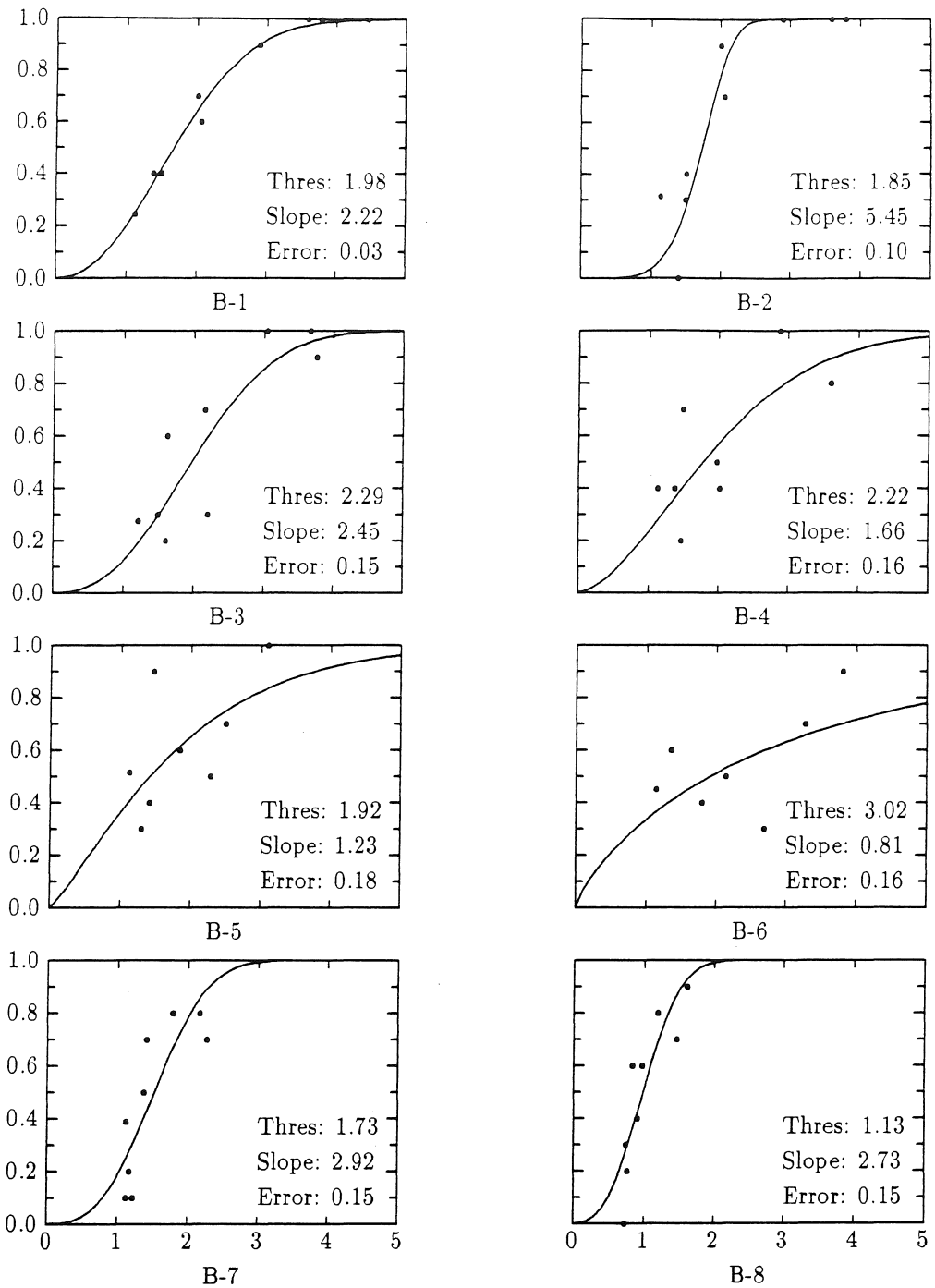


Figure B.1c. Psychometric functions with the *nonrigidity* discrimination measure as the independent variable. Measures were computed using 5% noise and the *actual* angular velocity; thresholds are absolute thresholds. (See column Nonrigid in Tables B-1 through B-8.)

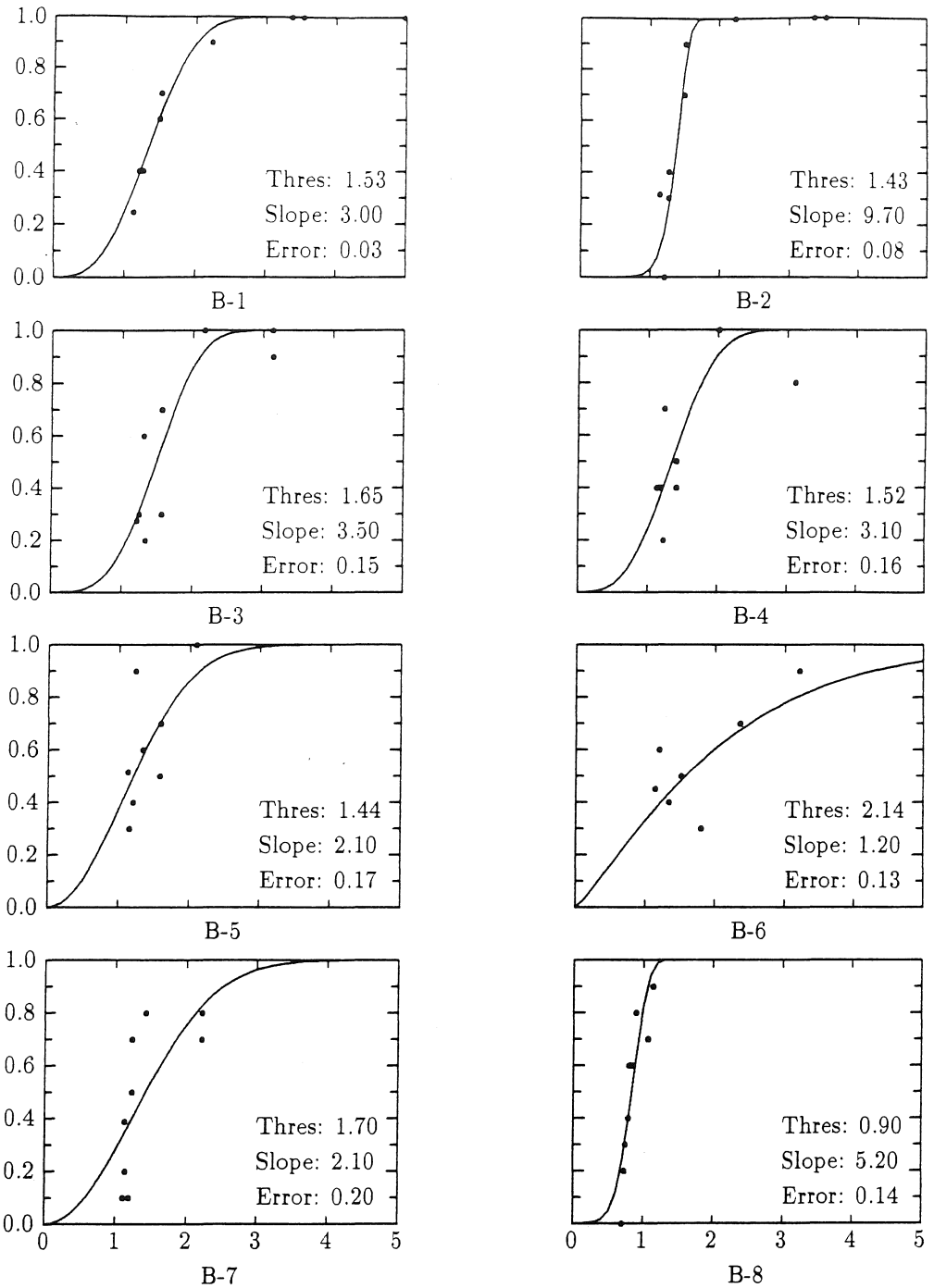


Figure B.1d. Psychometric functions with the *nonrigidity* discrimination measure as the independent variable. Measures were computed using 5% noise and the *estimated* angular velocity. Thresholds are absolute thresholds. (See column Nonrigid in Tables B-1 through B-8.)

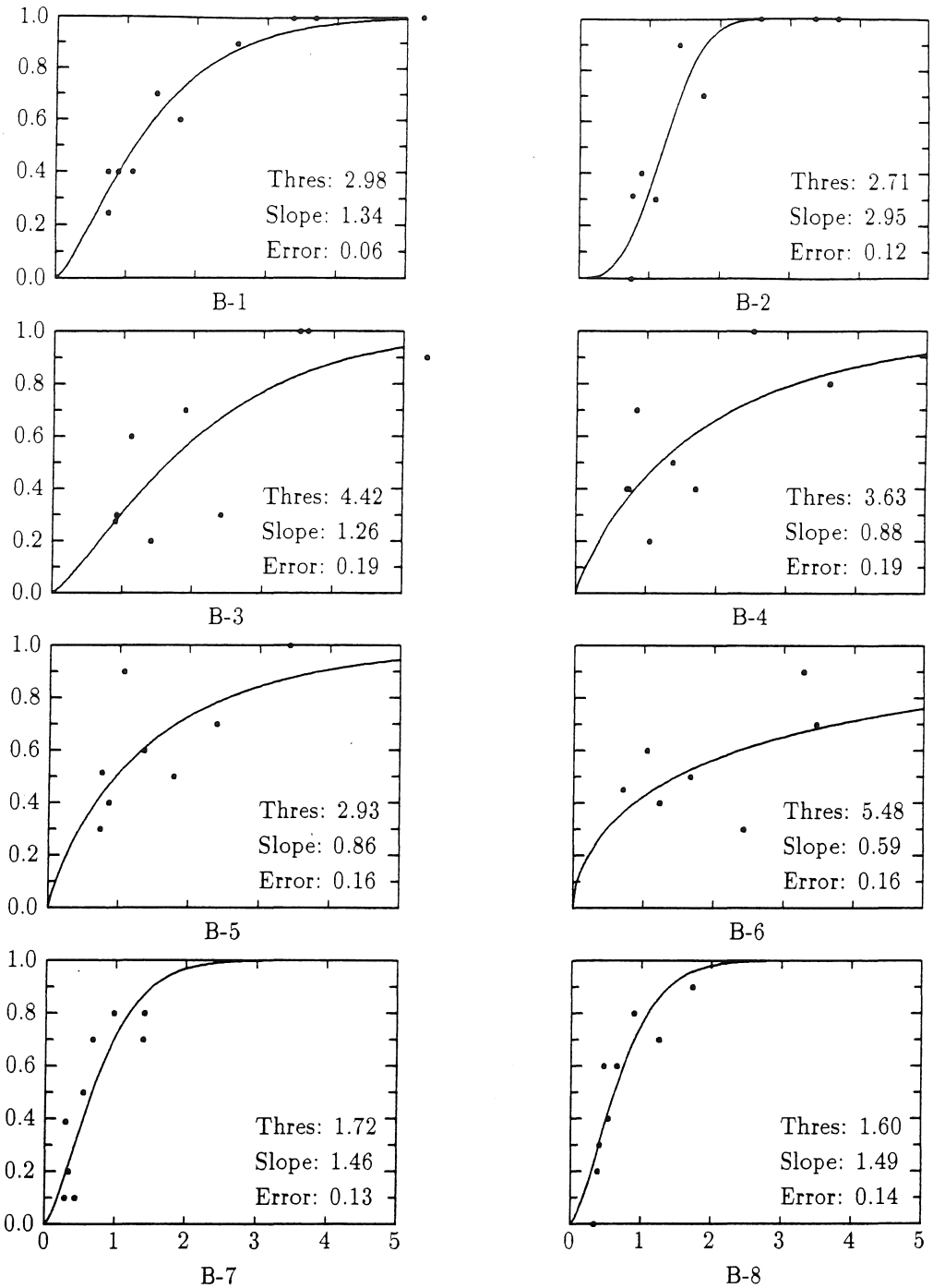


Figure B.1e. Psychometric functions with the *rim prediction* discrimination measure as the independent variable. Measures were computed using 5% noise and the *actual* angular velocity; thresholds are absolute thresholds. (See column Rim in Tables B-1 through B-8.)

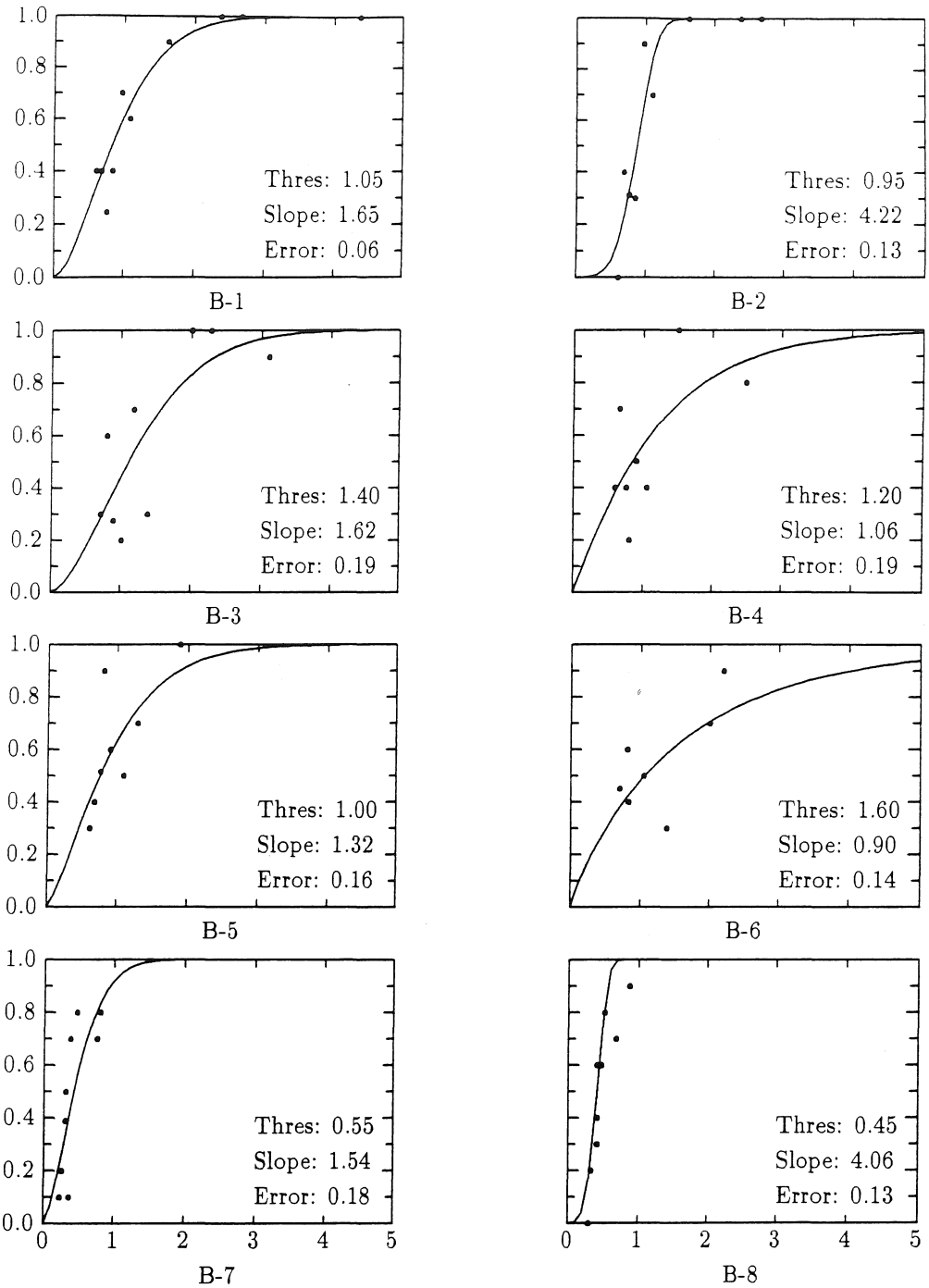


Figure B.1f. Psychometric functions with the *rim prediction* discrimination measure as the independent variable. Measures were computed using 5% noise and the *estimated* angular velocity. Thresholds are absolute thresholds. (See column Rim in Tables B-1 through B-8.)

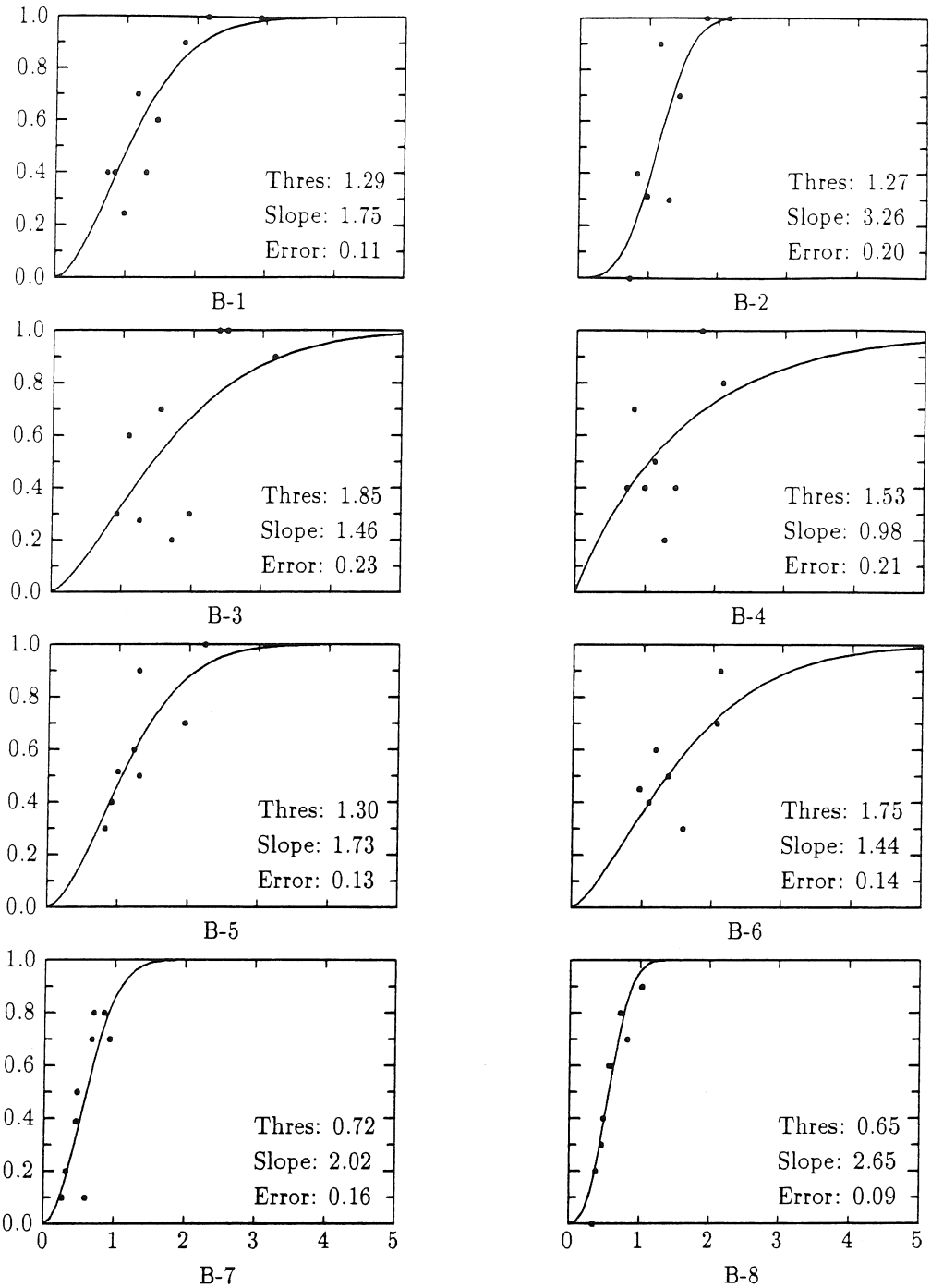


Figure B.1g. Psychometric functions with the *tangent plane prediction* discrimination measure as the independent variable. Measures were computed using 5% noise and the *actual* angular velocity. Thresholds are absolute thresholds. (See column Tangent in Tables B-1 through B-8.)

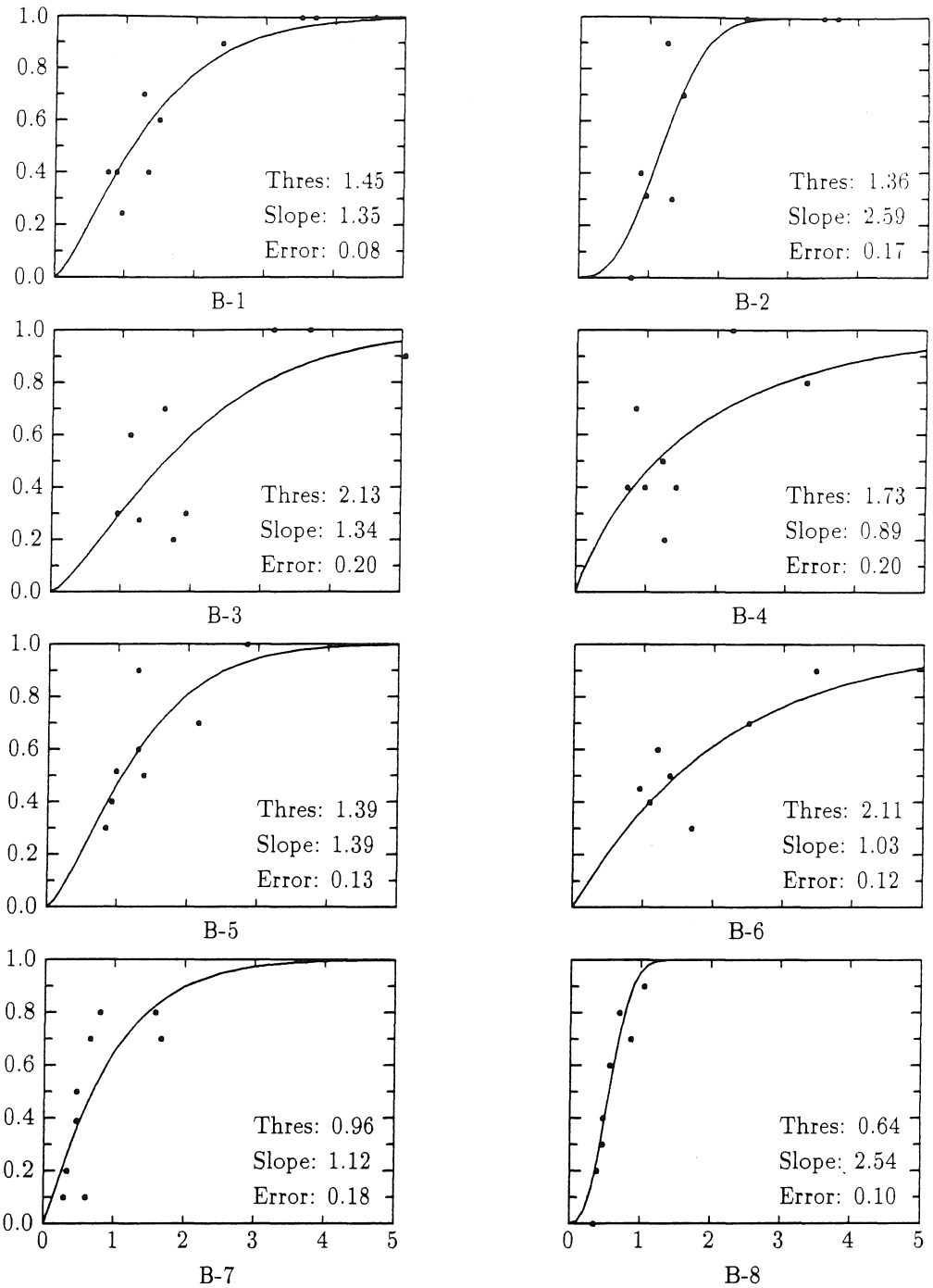


Figure B.1h. Psychometric functions with the *tangent plane prediction* discrimination measure as the independent variable. Measures were computed using 5% noise and the *estimated* angular velocity. Thresholds are absolute thresholds. (See column Tangent in Tables B-1 through B-8.)

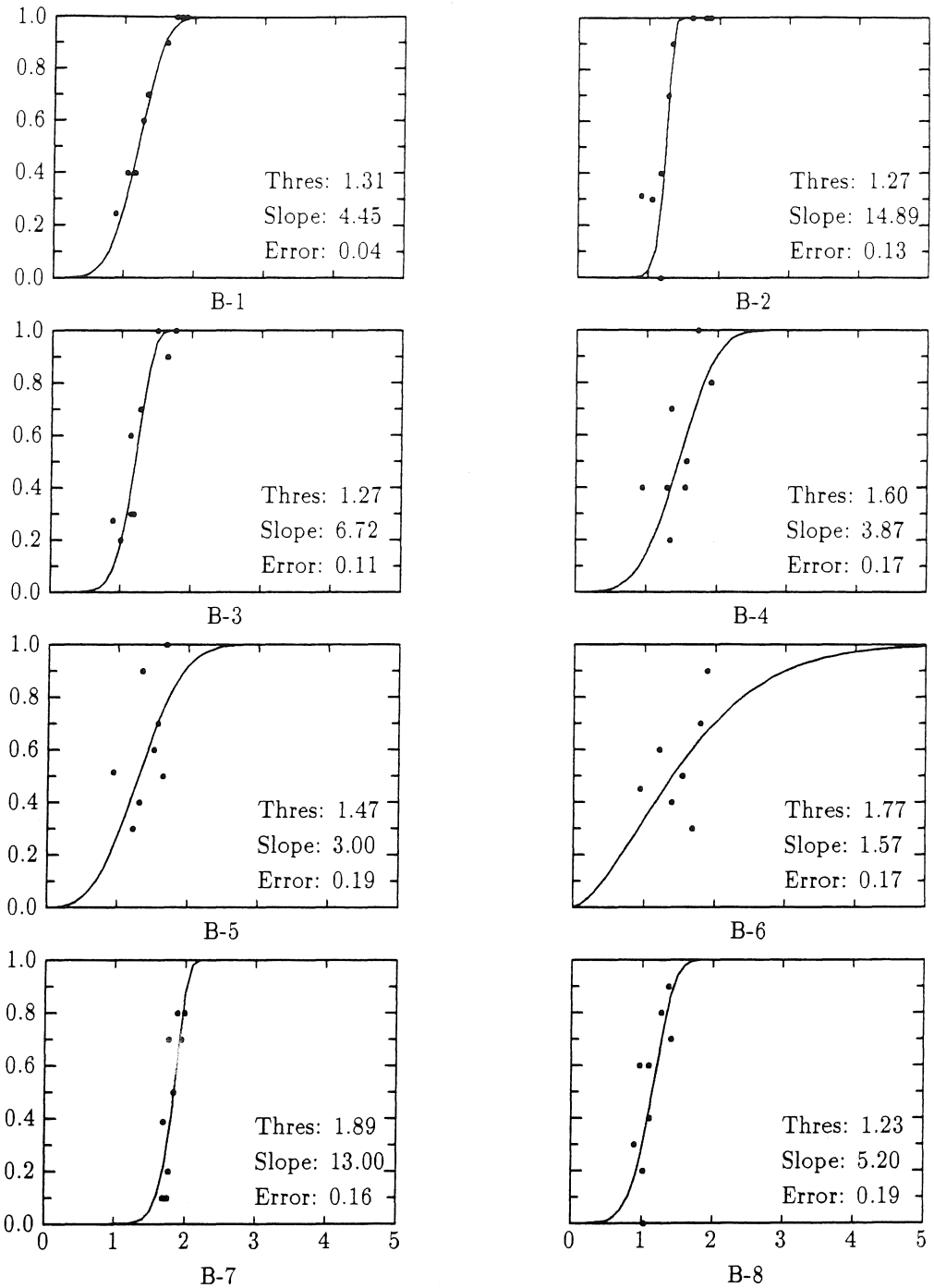


Figure B.1i. Psychometric functions with the *contraction* discrimination measure as the independent variable. Measures were computed using 5% noise. Thresholds are absolute thresholds. (See column Contr in Tables B-1 through B-8.)

Appendix C

Glossary

Aspect. A qualitative description of a view of a solid object. This description is stable under small changes in vantage point; visual events change the aspect (section 2.8, Figure 2.13a).

Aspect graph. See Visual potential.

Asymptotic directions. The directions at a surface point in which normal curvature is zero. There are two such directions for hyperbolic points (Figure C.3), and only one for parabolic points (Figure C.1). At planar points, all directions are asymptotic, and there are no asymptotic directions at elliptic points (Figure C.2).

Asymptotic line. A line on a surface which is everywhere tangent to an asymptotic direction.

Bitangent. A straight line tangent to two successive contours (section 4.1 and Figure 4.8).

Concave point. A point whose principal curvatures both are negative. Its Gaussian curvature is positive; its mean curvature negative. Points in a dent are concave. See Figure C.2.

Conjugate directions. Let $\delta u : \delta v$ and $du : dv$ be the direction numbers of two directions at a point on a surface. The two directions are conjugate iff $(d\mathbf{x} \cdot \delta\mathbf{N}) = 0$, where $d\mathbf{x} = \mathbf{x}_u du + \mathbf{x}_v dv$ and $\delta\mathbf{N} = \mathbf{N}_u \delta u + \mathbf{N}_v \delta v$. Asymptotic directions are self-conjugate. Principal directions are conjugate. Figure 4.2 shows the conjugate directions in the neighborhood of an elliptic and a hyperbolic point.

Contour. See occluding contour.

Convex point. A point whose principal curvatures both are positive. Its Gaussian and mean curvatures are positive. All points on an ellipsoid are convex. See Figure C.2.

Curvature model. A representation of solid shape based on measures of surface curvature such as Gaussian curvature or mean curvature (Figure 4.10 and section 6.1).

Cusp. (i) One of the two stable singularities of mappings of the type $\gamma : \Re^2 \rightarrow \Re^2$ (Whitney, 1955); we considered singularities of the visual mapping and the Gauss mapping. At a cusp, γ is locally 3-to-1. Cusps occur along folds. The Gauss map has a cusp at pedal points. (ii) A point on an occluding contour at which the viewing direction is in an asymptotic direction, causing the contour to stop (special case of i).

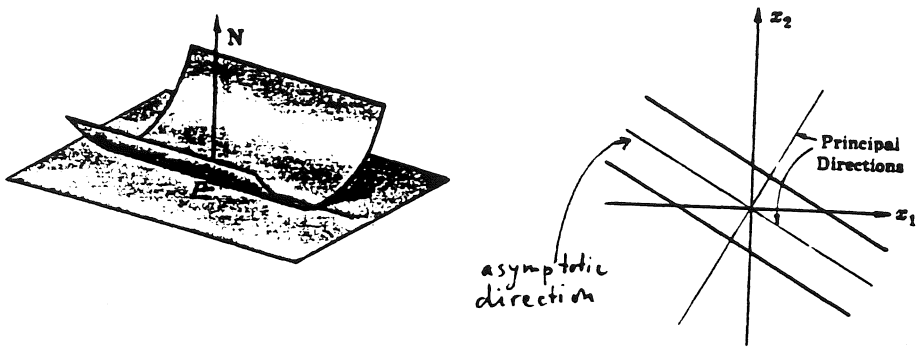


Figure C.1. Parabolic point (from Lipschutz, 1969).

Cylindrical point. See Parabolic point and Figure C.1.

Distance function. A function $\rho(\mathbf{x})$ defined for points $\mathbf{x}$ in the image indicating the distance to the nearest surface point in the direction specified by $\mathbf{x}$.

Dupin's indicatrix. Imagine the tangent plane at some surface point P; now cut the surface by planes that are parallel to the tangent plane. This cross section yields Dupin's indicatrix (see Figures C.1, C.2, and C.3). Convex and concave points yield elliptic indicatrices; saddle-shaped or hyperbolic points yield hyperbolic ones. The indicatrix of parabolic points (also known as cylindrical points) consists of straight lines. The indicatrix can also be obtained by following conjugate directions in the neighborhood of P (Figure 4.1).

Elliptic point. A point whose principal curvatures have equal sign (Figure C.2); thus, its Gaussian curvature is positive. The point is either convex or concave.

Equidistance lines. Lines on a surface along which the distance to the observer remains constant (section 2.5, Figures 2.6 and 2.8).

Fixed point. A point P_0 on an occluding contour C_0 which corresponds to more than one point P_t on contours C_t, where C_t is a family of contours of some object. As far as rotation is concerned, there are fixed points if and only if an object rotates about a frontoparallel axis or about the viewing direction itself (assuming orthographic projection throughout). See section 4.1.

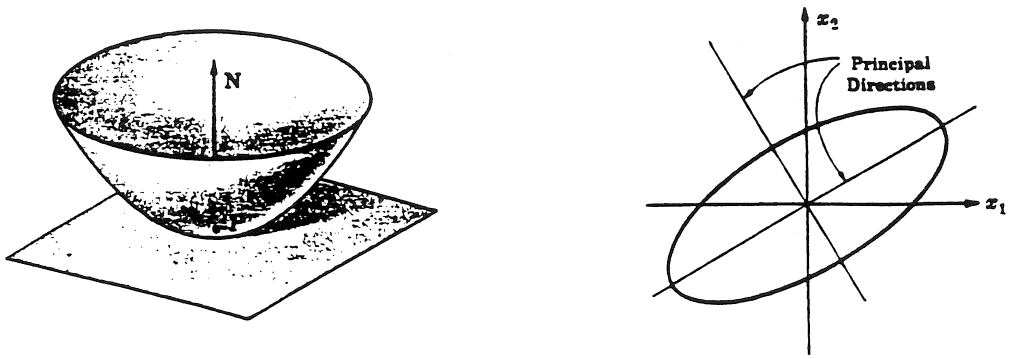


Figure C.2. Elliptic point (from Lipschutz, 1969).

Flecnodal point. An inflection point of an asymptotic line; a flecnodal line consists of flecnodal points. The swallowtail visual event occurs when the viewing direction is tangent to a flecnodal line (Figure 2.1).

Focus of expansion. See Focus of radial outflow.

Focus of radial outflow. The optic flow field for an observer translating through the environment consists of vectors that all point toward or away from a single point in the image, the so-called focus of radial expansion (section 2.5). Gibson had originally called it the focus of expansion but this is, technically, a misnomer as expansion is actually zero at the FRO (Koenderink & van Doorn, 1981).

Fold. (i) One of the two stable singularities of mappings $\gamma : \Re^2 \rightarrow \Re^2$ (Whitney, 1955); we considered singularities of the visual mapping and the Gauss mapping. Along a fold, γ is 2-to-1. At fold points of the visual mapping, the surface normal is perpendicular to the viewing direction. The fold loci of the Gauss mapping are parabolic curves, and vice versa.

FRO. See Focus of radial outflow.

Gauss map. Also known as the spherical mapping or mapping by parallel normals. Let $N : M \rightarrow S^2$ be the Gauss map, where M is a 2D manifold and S^2 the unit or Gauss sphere. For $x \in M, s \in S^2, N(x) = s$ and $N(x)$ is the surface normal at x. In other words, the Gaussian image of a surface point x is the point s on the Gauss sphere whose direction is that of the surface normal at x; hence the name mapping by parallel normals. See Figures 2.3 and 2.4 for examples of Gaussian images of solid objects.

Gaussian curvature. The product of the two principal curvatures. It is the determinant of the Gauss map.

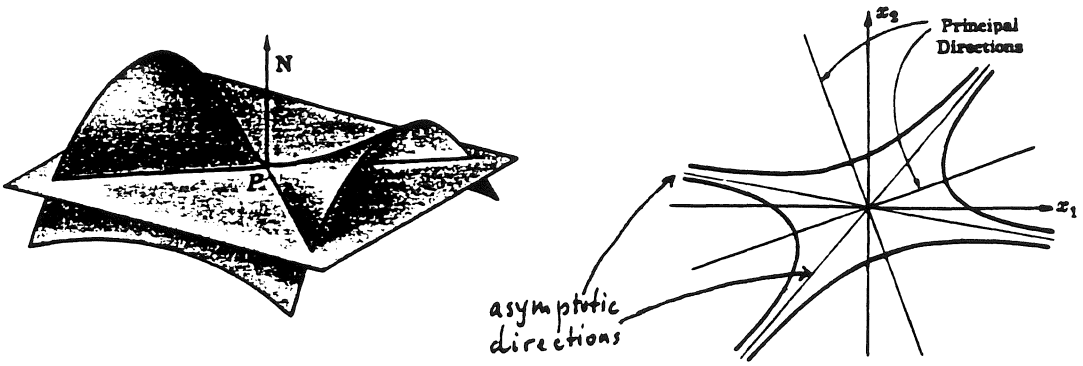


Figure C.3. Hyperbolic point (from Lipschutz, 1969).

Hyperbolic point. A point with one negative and one positive principal curvature. The surface has locally the shape of a saddle. It has negative Gaussian curvature. See Figure C.3.

Isophote. A line on the surface of an object along whose image luminance is constant.

Level lines. See Equidistance lines.

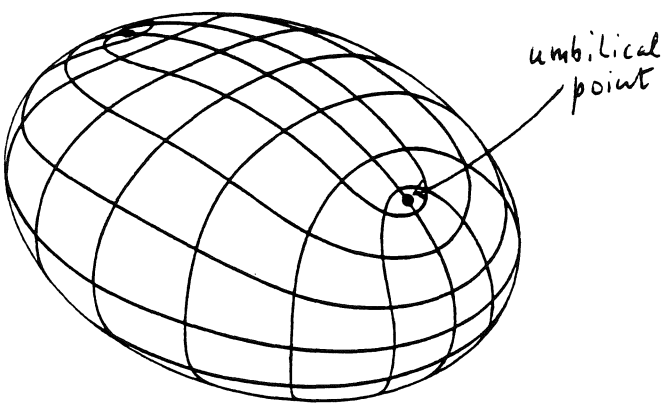


Figure C.4. Lines of curvature (from Hilbert & Cohn-Vossen, 1983).

Lines of curvature. Lines on a surface that are everywhere tangent to a principal direction. Because the principal directions are orthogonal everywhere except at umbilical points, the lines of curvature form an orthogonal net (except at umbilical points). See Figure C.4 for the lines of curvature on an ellipsoid; an ellipsoid has four umbilical points, two of which are visible in Figure C.4.

Mean curvature. The average of the two principal curvatures at a point.

Nonrigidity measure. A measure of the nonrigidity of a collection of points developed by Braunstein, Hoffman, & Pollick (1989). The measure is the mean of the variance of all interpoint distances (section 6.1).

Normal curvature. The curvature of a normal section at a surface point.

Normal section. A section of a surface at a point P by a plane that contains the local surface normal at P.

Occluding contour. The projection of a rim; the silhouette is an example of an occluding contour.

Parabolic point. A point one of whose principal curvatures is zero; the surface is locally like a cylinder whose axis is in the direction of zero principal curvature (they are sometimes called cylindrical points). See Figure C.1. The Gaussian curvature of a parabolic point is zero. Parabolic curves are the folds of the Gauss map (Figures 2.3 and 2.4).

Pedal point. A point on a parabolic curve where the direction of the parabolic curve equals the direction of zero principal curvature. At pedal points, the Gauss map has a cusp (Figure 2.4).

Planar point. A point whose principal curvatures are both zero.

Point of bitangency (POB). A point on an occluding contour whose tangent is also tangent to another point on another contour (Figure 4.8).

Point of correspondence (POC). A point on an occluding contour which has the same pre-image as some point on another contour of the object (section 4.1).

Principal curvature. The normal curvature in a principal direction.

Principal directions. At a point on a smooth surface there are two directions in which normal curvature is extremal, the so-called principal directions. The principal directions are perpendicular, except at umbilical points where every direction is a principal direction. See Figures C.1, C.2, and C.3.

Radial curvature. Radial curvature at a point on the rim refers to the normal curvature obtained by the normal section containing the viewing direction.

Rim. A curve on a smooth solid object along which the local surface normal is perpendicular to the local viewing direction. Rims are the fold loci of the visual mapping. Rims separate visible from invisible points on opaque objects.

Rim prediction model. A representation of solid shape in which the current rim is used to predict future rims (section 6.1).

Slant angle. (i) The slant angle σ of a vector such as a surface normal or axis of rotation is the angle between the image plane and that vector. (ii) $\sigma = \text{atan}(|\mathbf{F}|)$, where $\mathbf{F}$ is the slant field.

Slant field. $\mathbf{F} = \text{grad}\log(\rho_0/\rho)$, where ρ is the distance function and ρ_0 is an arbitrary constant (Koenderink, 1986). $\mathbf{F}$ is a 2D vector in the image in the direction of largest local decrease in distance, whose length is proportional to this local decrease. Along occluding contours, $|\mathbf{F}| = \infty$.

Spherical mapping. See Gauss mapping.

Tangent plane prediction model. Same as rim prediction model, but using the tangent plane along a rim in addition to the rim itself.

Tilt. The angle between the image of a vector and some particular reference direction in the image such as the vertical.

Transverse curvature. Curvature of the occluding contour.

Umbilical point. A point at which normal curvature is the same in all directions. Examples are points on a plane (normal curvature is zero in all directions) and points on a sphere (normal curvature is the same but nonzero in all directions). At an umbilical point all directions are principal directions. Figure C.4 shows two of the four umbilical points on an ellipsoid.

Visual event. A qualitative change in the image of an object, usually in the context of occluding contours, brought about by a change in vantage point. Figure 2.1 shows all the generic visual events for occluding contours.

Visual potential. The graph of all aspects of an object. Each vertex of the graph represents a particular aspect and with it the set of nearby vantage points from which the object has this aspect; the edges connecting the vertices represent the observer motion and the associated visual event which separate pairs of aspects (see Figure 2.13b).